DigItalia

DigItalia

ISSN 1972-6201 Anno X, Numero 1/2 - **2015**

RIVISTA DEL DIGITALE NEI BENI CULTURALI

ICCU-ROMA

Anno X, Numero 1/2 - **2015**

SISTEMA MODUS

DigItalia

Anno X, Numero 1/2 - **2015**

Rivista del digitale nei beni culturali

ICCU-ROMA

SISTEMA MODUS

ICCU

Istituto centrale per il catalogo unico
delle biblioteche italiane e per le informazioni bibliografiche
http://www.iccu.sbn.it

Copyright © ICCU - Roma

La riproduzione totale o parziale del contenuto della rivista
è ammessa con obbligo di citazione

SISTEMA MODUS

DigItalia

Rivista del digitale nei beni culturali
ISSN 1972-6201
Anno X, Numero 1/2 - Dicembre 2015

In copertina:
L'immagine è una libera elaborazione grafica della testa della statua di Apollo del I sec. d.c. (Civitavecchia, Museo Nazionale), copia da un originale greco avvicinabile all'Apollo di Leochares (IV sec. a.c.)

Direttore Fondatore
Marco Paoli

Direttore Responsabile
Simonetta Buttò

Comitato di Redazione
Simonetta Buttò
Lucia Basile
Silvana de Capua
Maria Delfino
Carla Di Loreto
Maria Cristina Di Martino
Valeria Di Vita Cafasso
Vilma Gidaro
Maria Cristina Mataloni
Massimo Menna
Lucia Negrini
Patrizia Nuccetelli
Paola Puglisi
Vittoria Tola

Grafica & Impaginazione
MLA&Partner - Roberta Micchi

Produzione e Stampa
Istituto Poligrafico e Zecca dello Stato S.p.A.
Roma

Editore
ICCU
Istituto centrale per il catalogo unico
delle biblioteche italiane
e per le informazioni bibliografiche
Viale Castro Pretorio, 105
00185 Roma
T +39 06 49.210.425
F +39 06 49.59.302

http://digitalia.sbn.it

In attesa di registrazione al Tribunale di Roma

SISTEMA MODUS

ICCU

Comitato Scientifico

Osvaldo Avallone
Giovanni Bergamin
Sandro Bulgarelli
Simonetta Buttò
Rossella Caffo
Rosaria Campioni
Gisella Capponi
Laura Ciancio
Flavia Cristiano
Gianfranco Crupi
Andrea De Pasquale
Maria Cristina Di Martino
Pierluigi Feliciati
Antonia Ida Fontana
Paolo Galluzzi
Marina Giannetto
Maria Guercio
Mauro Guerrini
Klaus Kempf
Maria Vittoria Marini Clarelli
Patrizia Martini
Maurizio Messina
Maria Cristina Misiti
Laura Moro
Maria Teresa Natale
Patrizia Nuccetelli
Marco Paoli
Alberto Petrucciani
Massimo Pistacchi
Marco Pizzo
Paola Puglisi
Antonia Pasqua Recchia
Gino Roncaglia
Don Valerio Pennasso
Marco Santoro
Maria Letizia Sebastiani
Rosaria Maria Servello
Giovanni Solimine
Maria Carla Cavagnis Sotgiu
Laura Tallandini
Anna Maria Tammaro
Costantino Thanos
Vittoria Tola
Paul Weston

SISTEMA MODUS

SOMMARIO

dicembre 2015

SAGGI

Sviluppi internazionali dei linked open data (LOD) nelle biblioteche: sfide ed opportunità 9
di Anna Maria Tammaro

Biblioteche digitali e Linked Open Data: contenuti, connessioni, link, argomentazioni 20
di Maurizio Lana

Web dei dati alla Biblioteca nazionale centrale di Firenze 30
di Anna Lucarelli

L'esperienza LOD (Linked Open Data) del Museo Galileo 40
di Stefano Casati, Tiziana Possemato

Nuove prospettive per il Web archiving: gli standard ISO 28500 (formato WARC) e ISO/TR 14873 sulla qualità del Web archiving 49
di Stefano Allegrezza

PROGETTI

Linked data: il mondo di internet e il ruolo delle biblioteche, degli archivi e dei musei 65
di Margherita Aste, Maria Cristina Mataloni, Luca Martinelli

Il progetto Clavius on the Web entra nelle scuole 74
di Andrea Marchetti, Simone Marchi, Martín Maria Morales

I possessori nella Biblioteca nazionale di Napoli: un Archivio di immagini 85
di Simona Pignalosa

SISTEMA MODUS

Mostre digitali:
due risultati del progetto AthenaPlus 96
di Maria Teresa Natale, Giuliana De Francesco

Il Polo digitale degli istituti culturali di Napoli 102
di Luigi Cerullo

La banca dati del manifesto politico 121
di Siriana Suprani

EVENTI

Beni culturali ecclesiastici e comunicazione.
Un umanesimo digitale possibile? 131
di Don Valerio Pennasso

Convegno Anno Europeo 2015
"Gli Obiettivi di Sviluppo sostenibile dell'Onu
e la Carta di Milano" 134
di Massimina Cattari

SEGNALAZIONI

Il progetto Europeana Food and Drink
Il cibo, il bere e il patrimonio culturale 141
di Elisa Sciotti

Saggi

Sviluppi internazionali dei linked open data (LOD) nelle biblioteche: sfide ed opportunità

Anna Maria Tammaro
IFLA Sezione Library Theory and Research

Le biblioteche sono spinte dalla tecnologia Linked Open Data (LOD) a strutturare il mondo bibliografico che le caratterizza in quello che è il Web dei dati. Questa spinta ad adottare i LOD rappresenta per le biblioteche una sfida che non è solo tecnologica: le biblioteche stanno sperimentando nel Web dei dati attività insolite come la connessione (il collegamento tra dati), la condivisione (la convergenza dei dati di diverse istituzioni), la partecipazione (nuove partnership con gli utenti) per nuovi servizi. Le promesse di miglioramento di servizio agli utenti devono tuttavia essere ancora valutate. Occorre fare ricerca sull'impatto dei LOD sul miglioramento del servizio agli utenti e sulle conseguenze per le procedure interne delle biblioteche.

1. Standard e mappatura

Il termine Linked Open Data (LOD) definisce la pubblicazione di risorse digitali che garantiscono stabilità di accesso ai dati ed al loro contenuto intellettuale, insieme alla loro ri-usabilità a livello granulare[1]. Tim Berners Lee ha lanciato il concetto degli Open Data nel 2007 descrivendo come si potesse migliorare il Web pubblicando e collegando dati strutturati. Mentre l'attuale Web consente di accedere ai documenti (pagine Web), il Web dei dati (chiamato dal W3C "Web semantico") ha l'ambizione di consentire l'accesso, sia a persone che ai computer, ai dati. RDF (Resource Description Framework) sta al Web dei dati come HTML sta al Web dei documenti. Nella terminologia del W3C, RDF consente di descrivere in modo non ambiguo le "cose" del mondo reale e le loro relazioni, non solo quindi le risorse digitali, ma gli oggetti reali ed i concetti astratti a cui le risorse si riferiscono.

[1] Gli Open Data sono associati con una nuova generazione del Web chiamato dal Consorzio W3C "semantico" e caratterizzato da:
- un framework comune dei dati (RDF) che, invece di limitarsi allo scambio di documenti, consente di estrarre i dati da diversi depositi e applicazioni
- un vocabolario condiviso (ontologie) per connettere i dati tra di loro e agli oggetti che rappresentano, che rende possibile la navigazione per "significato" e non solo per collegamento ipertestuale (link).

SISTEMA MODUS

1.1 LOD nelle biblioteche Linked Open Data

La sfida e l'opportunità che si offre ora alle biblioteche è quella di pubblicare i loro dati (metadati e contenuti) come Linked Open Data nel Web dei dati. Tra tutte le istituzioni che pubblicano i dati aperti, le biblioteche sembrano quelle che dovrebbero fare lo sforzo minore per ri-utilizzare i loro dati, che nascono originariamente già strutturati. Il primo Gruppo di lavoro sulle biblioteche (Library Linked Open Data) è stato avviato proprio dal Consorzio W3C (2011) che ha predisposto un rapporto sui primi progetti di biblioteche.

Nel mondo delle biblioteche, IFLA è stata la prima ad interessarsi ai Library Linked Open Data. La Sezione IFLA Information Technology ha organizzato un primo Satellite sul Web semantico a Firenze nel 2009, durante il Convegno di Milano, dal titolo "*Emerging trends in technology: libraries between Web 2.0, semantic Web and search technology*", che ha avuto un grande successo di pubblico. La Sezione IFLA Information Technology ha quindi riproposto il Satellite coinvolgendo altre Sezioni interessate (la Sezione Knowledge management, la Sezione Cataloguing e la Sezione Classification and Indexing) nel Convegno di Göteborg (2010). Questa seconda sessione ha stimolato l'idea di creare uno Special Interest Group (SIG) Semantic Web per raccogliere la comunità degli sviluppatori di progetti LOD nelle biblioteche.

Un'altra comunità internazionale è LOD-LAM, interessata alla convergenza di archivi, biblioteche e musei (LAM Library Archives Museum), che dal 2010 si è sviluppata aggregandosi intorno ad un sito Web[2] e a un gruppo di discussione, e che organizza ogni anno un summit in un'Area geografica diversa.

I primi progetti LOD nelle biblioteche hanno portato a riflettere su varie problematiche e possono illustrare molte fasi critiche del processo di passaggio dalla progettazione alla realizzazione della pubblicazione di LOD. Per analizzare le opportunità e le sfide delle biblioteche nel Web dei dati, è stato interessante per me leggere alcune delle iniziative del gruppo LOD-LAM, le presentazioni alle Conferenze Satellite del SIG dell'IFLA Semantic Web, i risultati del Library Linked Data Group, i risultati di alcuni dei progetti come Learning Linked Data[3], Linked Open Vocabularies[4], nonché leggere alcune delle relazioni dei Convegni internazionali, a cominciare da quello tenutosi a Firenze *Global Interoperability and Linked Data in Libraries* nel 2012[5]. Il sito personale di Karen Coyle[6] è un'altra delle

[2] http:/lodlam.net.

[3] Progetto finanziato da ILMS nel 2012, <http://lld.ischool.uw.edu/wp/>.

[4] http://lov.okfn.org/dataset/lov/.

[5] Il 18 e 19 giugno 2012 si è svolto, nell'Aula Magna dell'Università di Firenze, il seminario Global interoperability and linked data in libraries, promosso da Università di Firenze, Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche (ICCU), Biblioteca nazionale centrale di Firenze (BNCF), Casalini Libri, Comune di Firenze, Conferenza dei rettori delle università italiane (CRUI), Associazione italiana biblioteche (AIB), Istituto di teoria e tecniche dell'informazione giuridica del Consiglio nazionale delle ricerche (ITTIG-CNR), Fondazione Rinascimento digitale, Associazione italiana editori (AIE), in collaborazione con la rivista JLIS.it (www.jlis.it).

[6] http://kcoyle.blogspot.it.

SISTEMA MODUS

risorse che ho utilizzato per evidenziare gli sviluppi internazionali dei LOD nelle biblioteche.

Le comunità bibliotecarie si sono concentrate inizialmente sulla "mappatura" tra gli standard bibliotecari e gli standard dei LOD, anche se biblioteche diverse (per esempio la British Library e la Bibliothèque nationale de France) hanno scelto diverse combinazioni di standard e vocabolari. Gli standard che IFLA sta sviluppando in questo settore, come FRBR, RDA e la maggior parte delle norme IFLA, sono ora disponibili come Linked Data. Tuttavia, le norme IFLA non sono ancora utilizzate dalla maggioranza delle biblioteche, e i vari gruppi di discussione si sono concentrati sul perché e su come migliorare la situazione presente. La Library of Congress si è assunta il compito di guidare la progettazione di dati aperti collegata al MARC, il formato del catalogo di biblioteca più diffuso, vecchio di mezzo secolo. Il risultato è stato BIBFRAME, un framework bibliografico che ha lo scopo di fornire i concetti di base per i catalogatori, integrando gli standard di Internet ai dati della biblioteca in modo efficace ed efficiente. In Europa, Europeana ha avuto un ruolo guida per i LOD nelle biblioteche, affrontando diverse problematiche e ha pubblicato un video[7] per spiegare cosa sono i LOD e perché apportano benefici sia agli utenti che ai creatori di dati.

Coyle (2014), durante la sua partecipazione al Convegno AIUCD a Bologna, ha ribadito con forza che il comportamento di ricerca degli utenti è cambiato e la ricerca dell'informazione parte ora dal Web; da qui l'urgenza di strutturare il mondo bibliografico in quello che è il Web dei dati. Questa spinta ad adottare i LOD rappresenta per le biblioteche una sfida che non è solo tecnologica: le biblioteche devono pensare in "grande" per diventare parte del Web dei dati, inteso come eco-sistema digitale. Per evidenziare la portata del cambiamento, Health e Bizer (2011) definiscono i LOD per le biblioteche come il passaggio da "global data islands" a "global data space": le biblioteche stanno sperimentando nel Web dei dati attività per loro insolite come la connessione (il collegamento tra dati), la condivisione (la convergenza dei dati di diverse istituzioni), la partecipazione (nuove partnership con gli utenti) per nuovi servizi.

I LOD sono quindi un'opportunità per le biblioteche ed è più che mai necessario costruire una comunità LOD nelle biblioteche, ma bisogna aver consapevolezza della sfida che rappresentano per le biblioteche, che hanno avuto finora un approccio chiuso, descritto in letteratura come quello di "sylos" o come "walled garden".

2. Le biblioteche nei Linked Open Data

L'indagine compiuta dal Consorzio W3C per i Library Linked Open Data ha evidenziato che le biblioteche hanno pubblicato finora le seguenti tipologie di dati (Fig. 1): dati bibliografici, liste di autorità, vocabolari, dati d'archivio, oggetti digitali, dati sulle collezioni, dati sociali ed usi.

[7] Linked Open Data. What is it?,< http://vimeo.com/36752317>.

SISTEMA MODUS

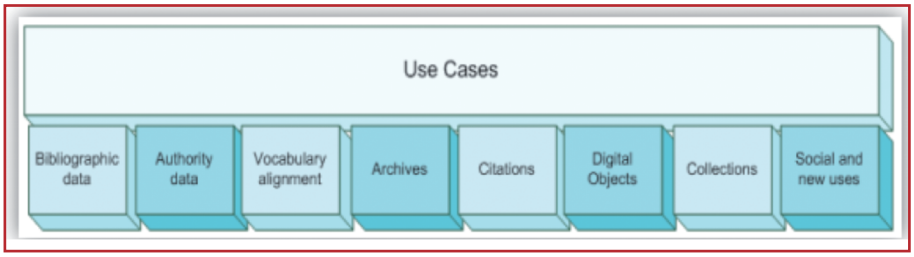


Figura 1: *W3C Library Linked Data report*

La prima biblioteca a pubblicare LOD è stata la Biblioteca Nazionale di Svezia nel 2008, seguita dalle biblioteche nazionali di Ungheria, Germania, Francia e la Library of Congress. Dopo le biblioteche nazionali hanno iniziato a pubblicare LOD le reti bibliotecarie come OCLC. Il contributo di OCLC alla visibilità dei LOD di biblioteche è stato importante e ha prodotto modelli di dati e insiemi di dati RDF per molti cataloghi di biblioteche, tra cui WorldCat, WorldCat Opere, liste di autorità come VIAF, vocabolari come FAST (Faceted Application of Subject Terminology) e la Classificazione Decimale Dewey. Tutti questi dati sono codificati nella Schema.org, il vocabolario accettato dai principali motori di ricerca di tutto il mondo per gli indici e la visualizzazione di dati strutturati. Insieme alle biblioteche nazionali ed ai servizi bibliografici hanno pubblicato LOD anche alcune biblioteche speciali come ad esempio la biblioteca della FAO.

2.1 Dati connessi

Il primo obiettivo perseguito con la pubblicazione dei LOD è stato quello di rendere i servizi in linea delle biblioteche più interessanti e facilitare la ricerca. Le biblioteche hanno quindi usato la loro collezione di dati per attirare gli utenti agli OPAC, mettendo i motori di ricerca in grado di connettersi ai cataloghi e rendendo molto più facile agli utenti trovare parole chiave, libri, informazioni, ecc.
Con questa prima connessione al Web dei dati, le biblioteche, forse senza piena consapevolezza, si sono assunte la responsabilità di "connettere" le loro collezioni con il mondo del Web. Tuttavia, il concetto di Web semantico è ancora nuovo e confuso per la comunità bibliotecaria e ci sono molte problematiche non tecniche che devono essere risolte. Per passare da una fase di prototipo, come molti degli attuali progetti delle biblioteche, ad una fase stabile di pubblicazione sistematica dei LOD, tutti gli insiemi di dati che sono stati pubblicati in Web dovrebbero essere connessi a dati sia interni alle biblioteche che esterni ad altre risorse nel Web. Tuttavia i dati delle biblioteche non sono sempre ri-usabili.

2.2 Dati aperti

I LOD sono dati che devono garantire la ri-usabilità: un pre-requisito è che i dati

siano "aperti". Ma come vanno considerati i dati bibliografici? Sono semplici fatti? Oppure sono opere dell'ingegno protette da copyright?
Il SIG di IFLA sul Web semantico ha chiarito la differenza tra Linked Data e Open Data, affermando che Linked Data è una tecnologia e Open Data è una politica, quindi ha senso combinare le due cose ed appoggiare decisamente l'apertura dei dati. È convinzione di IFLA che il massimo vantaggio dei dati nel Web venga dall'adottare una licenza "aperta", sostenendo l'appoggio ai principi per i dati bibliografici aperti[8], dati cioè liberamente disponibili per l'uso e il riuso da parte di chiunque e per qualsiasi scopo.
La proprietà dei dati bibliografici è tuttavia spesso incerta, ad esempio nel caso di dati provenienti da fornitori commerciali, oppure di dati di diverse biblioteche che cooperano in consorzi e reti bibliotecarie. Un errore abbastanza diffuso nei progetti LOD di biblioteche, è stato quello di non includere i dati di provenienza e i dati di proprietà intellettuale nell'insieme dei dati che veniva pubblicato nel Web. Senza un "data provider agreement", ad esempio Europeana non ha potuto pubblicare i dati nel Web.
Il problema è ancora più complicato a livello internazionale: ci sono norme di protezione della proprietà intellettuale diverse nei diversi Stati, ad esempio per le banche dati, per il pubblico dominio, per le limitazioni in genere all'accesso ai dati. Le biblioteche che hanno avviato progetti LOD, hanno usato le licenze Creative Commons: ad esempio la British Library e la Deutsche Nationalbibliothek hanno adottato la licenza CC0, la Biblioteca di Stanford la licenza CC-BY. Ma non tutte le licenze Creative Commons sono adatte ai dati da ri-utilizzare. Nei progetti avviati spesso mancano politiche scritte che riguardano sia la modalità di pubblicazione dei dati che le modalità di accesso consentite a questi (Greenberg 2007).

3. LOD oltre le biblioteche

LOD è una tecnologia che stimola una nuova interoperabilità, basata sulla condivisione dei dati per costruire un *global data space*. La sfida che le biblioteche affrontano in questo caso è quella di dover dare fiducia ad istituzioni esterne, che pubblicano numerose tipologie di dati, che si riferiscono a persone, istituzioni, film, musica, luoghi, libri e altre pubblicazioni, comunità in linea e una massa crescente di dati di enti pubblici (Heath, Bizer 2011).

3.1 Arricchire i dati con collegamenti (link)

Le biblioteche hanno utilizzato finora un linguaggio chiuso, che non si poteva facilmente comprendere al di fuori del mondo delle biblioteche, ma ora lo scenario dei LOD spinge le biblioteche a favorire la convergenza con altri domini e altre co-

[8] Il Manifesto sui dati bibliografici aperti è accessibile a:< http://openbiblio.net/principles/it/>.

munità. I LOD delle biblioteche si arricchiscono con collegamenti esterni e possono essere condivisi coi dati di istituzioni affini, come quelli degli archivi e dei musei, e con dati di altre istituzioni, come ad esempio DBpedia, VIAF, Geonames.
I progetti esistenti di LOD di biblioteche sono limitati spesso ad uno o due insiemi di dati ben integrati, di solito dello stesso creatore, e non usano ancora altri dati, interni o esterni alle biblioteche. Il collegamento dei dati delle biblioteche coi dati provenienti da altre applicazioni Web rappresenta quindi un grande cambiamento culturale che è solo al suo inizio. Il problema è che manca un modo di integrare i LOD provenienti da molte risorse e occorre trovare degli accordi. Servono anche accordi su come i dati arricchiti dai collegamenti esterni vengano poi presentati nell'interfaccia agli utenti.
Il dibattito attuale sull'integrazione dei dati LOD delle biblioteche con altre risorse ruota soprattutto intorno a che cosa fare rispetto a Wikipedia: collegare o non collegare a Wikipedia i dati bibliografici? Esempi positivi di istituzioni che hanno scelto di farlo possono includere la Biblioteca Nazionale di Firenze insieme alla BnF con il prototipo Opencat, e la BBC.

4. LOD e utenti

La motivazione principale dei progetti di LOD di biblioteca è stata quella di migliorare i servizi agli utenti, come limitare l'overload d'informazione, o facilitare il recupero di dati estratti da una grande quantità di risorse. Quali servizi LOD sono stati resi disponibili? Quali sono stati i risultati finora?
I servizi dei progetti LOD di biblioteca si sono concentrati sulle tecnologie e sulle interfacce, come la possibilità di visualizzazione dei dati accessibili, o la ricerca avanzata con strumenti che consentono all'utente di sintetizzare ed elaborare i dati. Gran parte della ricerca sui LOD in biblioteca, come conseguenza di queste priorità, si è concentrata su una metodologia quantitativa, basata sul rendimento di sistemi informativi su grandi insiemi di dati. Non c'è ancora una valutazione dell'impatto dei servizi LOD sugli utenti.
Possiamo allora affermare che i servizi LOD di biblioteca sono servizi disintermediati? Le tecnologie dei dispositivi mobili e delle reti sociali stanno accelerando il cambiamento di comportamento degli utenti, consentendo loro di partecipare attivamente e in modo autonomo alle attività di raccolta, creazione, conservazione e interpretazione dei contenuti. La sfida per le biblioteche è quella di usare questo contesto per dar vita ad una nuova era, caratterizzata da una cultura partecipativa nelle istituzioni culturali. Le biblioteche sono stimolate dai progetti LOD a incoraggiare gli utenti a interagire attivamente con i contenuti nel Web e anche a connettersi socialmente uno con l'altro. Il cambio culturale dei LOD in questo caso spinge le biblioteche a dare agli utenti la possibilità di aggiungere informazione ai dati che usano, a classificare le risorse, annotare i documenti, a contribuire alla storia

dell'evoluzione di concetti e idee. La cultura partecipativa (Giacardi 2012) sta emergendo come cultura innovativa nei beni culturali, che si appropria e risponde all'esplosione di tecnologie che rendono possibile per l'utente medio archiviare, annotare, ri-usare contenuti multimediali, in nuovi modi creativi ed efficaci. È proprio la cultura partecipativa che spinge a trovare nei LOD uno stimolo a esplorare la pubblicazione di dati di biblioteche, archivi, musei, dove non ci si limiti a navigare tra i dati resi pubblici del patrimonio culturale, ma dove il pubblico partecipi attivamente alla creazione di contenuti (Jenkins 2006). Queste attività partecipate hanno un impatto profondo e di trasformazione sulla comunicazione del patrimonio culturale, allargando così la comprensione e l'esperienza culturale sia del singolo che delle comunità.

La tendenza attuale delle biblioteche è quindi quella di facilitare la partecipazione attiva degli utenti nei progetti di LOD. Il progetto Linked Data for Libraries (LD4L)[9], ad esempio, è un progetto di due anni finanziato da Mellon, che vede la collaborazione tra Cornell University Library, Harvard Library Innovation Lab, e Stanford University Library per creare un deposito di dati (Scholarly Resource Semantic Information Store) sia all'interno delle istituzioni sia attraverso una rete coordinata, per catturare il valore intellettuale delle risorse, che bibliotecari e studiosi descrivono, annotano, organizzano, selezionano e usano. I bibliotecari impegnati nel progetto hanno discusso se limitarsi a riutilizzare i metadati della biblioteca o se usare dati specializzati sul tema, abbandonando la classificazione della biblioteca. Alla fine hanno deciso di concentrarsi sui dati d'autorità della biblioteca riferentesi al soggetto specifico. Questo progetto rappresenta uno degli sforzi più ambiziosi per integrare i LOD nelle operazioni quotidiane della biblioteca e le esperienze e gli insegnamenti tratti per affrontare i LOD da progetti come questo, aiutano a colmare una lacuna importante nella comprensione delle biblioteche in transizione, per contribuire a completare il puzzle delle applicazioni LOD nella comunità bibliotecaria. Rendere dati di autorità disponibili come dati connessi, con una licenza "aperta", sembra essere il beneficio che le biblioteche possono portare alla comunità.

L'argomentazione del Web dei documenti si fa creando link che connettono "entità informative" che secondo il punto di vista di chi crea il "link" mostrano una lettura dei fenomeni, e questo implica una selezione qualitativa. Nel Web dei dati invece i LOD creano connessioni automatiche, creando un contesto di relazioni, ma non creano conoscenza in senso stretto, il che vuol dire possibilità di errore, possibilità di dissenso, in assenza di criteri formali di validità. I potenziali vantaggi dei LOD stimolano biblioteche e comunità di utenti a collaborare insieme per pubblicare dati, in modo da favorire una maggiore utilità "nel flusso" del Web dei dati, creando connessioni inaspettate, usando i dati per creare valore, aggiungendo si-

[9] https://wiki.duraspace.org/display/ld4l/Project+Proposal.

gnificato, attività che possono essere realizzate dalla collaborazione con gli utenti ed altre istituzioni, ma non può essere solo compito delle biblioteche.

5. Conclusioni

Le tendenze internazionali sui LOD delle biblioteche sono evidenziabili nelle esperienze dei progetti LOD in Italia. La Tavola rotonda organizzata a Torino dal Salone del libro *Biblioteche digitali verso il futuro: accesso ai linked open data*[10] (interventi pubblicati in questo fascicolo di Digitalia) intendeva illustrare le concrete esperienze di biblioteche digitali italiane che pubblicano i LOD. C'è molto interesse da parte di bibliotecari e di utenti finali a saperne di più sui LOD di biblioteche e l'incontro di Torino ha avuto molto successo di pubblico. Le relazioni dei relatori hanno evidenziato che le biblioteche digitali italiane più avanzate sono ora connesse al Web e condividono le loro risorse con istituzioni esterne alle biblioteche, in transizione ad un ecosistema di dati connessi. Ad oggi, la maggior parte degli sforzi delle biblioteche di partecipare al Web dei dati si sono concentrati nel migliorare le possibilità di ricerca delle risorse da parte degli utenti finali. Occorre tuttavia fare ricerca sull'impatto che la transizione ai LOD ha comportato, o anche valutare l'impatto su approcci innovativi dei LOD nel campo del lavoro bibliografico, a partire dalle operazioni quotidiane delle biblioteche.

Sarebbe ottimale ad esempio convertire tutte le operazioni di dati di biblioteche in un ecosistema nativo di triple? Oppure, avrebbe più senso mantenere un archivio dati basato su campi bibliografici, e con un programma pubblicare i dati come triple? Quale effetto avrebbe l'adozione di questo approccio sugli attuali software di gestione bibliotecaria? Sui fornitori di servizi? Sui flussi di lavoro interni? Quale impatto infine sulla formazione bibliotecaria?

Durante la Tavola rotonda di Torino una domanda ha riguardato il ruolo dei bibliotecari per i LOD. Non tutti i bibliotecari possono essere creatori di LOD, è stata la risposta dei relatori. Ogni bibliotecario potrà però aiutare gli utenti a identificare le risorse nel Web dei dati per la loro area di specializzazione. L'ostacolo in questo caso potrebbe essere rappresentato solo dalla mancanza di capacità dei bibliotecari, o dalla mancanza di familiarità con le tecnologie per svolgere questo tipo di attività in ambiente LOD.

[10] Tavola rotonda con Anna Lucarelli (Biblioteca nazionale centrale Firenze), Stefano Casati (Biblioteca digitale IMSS Galileo), Maurizio Lana (Università del Piemonte orientale) e Luca Martinelli (ICCU), coordinata da Anna Maria Tammaro.

SISTEMA MODUS

Libraries are pushed by the technology of Linked Open Data (LOD) to structure their data in the Web of data. This change is not only a technological challenge: libraries are experiencing unusual activities in the Web data as the connection (the connection among data), sharing (the convergence of data from different institutions), participation (new partnerships with users) for new services. The promises of an improvement of services to users need to be further evaluated. We need to foster research on the impact of LOD on improving the service to users and the consequences for the internal procedures of libraries as well.

L'ultima consultazione dei siti Web è avvenuta nel mese di dicembre 2015.

Riferimenti bibliografici

Berners Lee 2007 — Berners, Lee, Tim *Open Data* 2007 http://www.w3.org/DesignIssues/LinkedData

Byrne – Goddard 2010 — Byrne, Gillian; Lisa, Goddard. *The Strongest Link: Libraries and Linked Data*, «D-Lib Magazine», 16 (2010), n. 11-12, http://dlib.org/dlib/november10/byrne/11byrne.html

Coyle 2009 — Coyle, Karen. *Making Connections, "*Library Journal", 134, (2009) n.7, p. 44-47.

Coyle 2010 — Coyle, Karen. *Understanding the Semantic Web: Bibliographic Data and Metadata*. "Library Technology Reports" 46, no. 1 (January 2010).

Coyle 2012 — Coyle, Karen. *Library linked data an evolution.* "Italian Journal of Library and Information Science" (2012) http://leo.cilea.it/index.php/jlis/issue/view/368/showToc

Coyle 2014 — Coyle, Karen; Gianmaria, Silvello e Anna Maria Tammaro. *Comparing Methodologies: Linked Open Data and Digital Libraries Third AIUCD Annual Conference on Humanities and Their Methods in the Digital Ecosystem.* New York: Springer, 2014, n. 3 http://dl.acm.org/ citation.cfm?id=2802615

Giaccardi 2012 — Giaccardi, Elisa (a cura di). *Heritage and Social Media: Understanding heritage in a participatory culture*. New York Routledge 2012

Greenberg 2007 — Greenberg, Jane. *Advancing the semantic Web via library functions "*Cataloging & classification quarterly" 43 (2007), n. 3-4, p. 203-225

Guerrini-Possemato 2012 — Guerrini, Mauro; Tiziana, Possemato. *Linked data: un nuovo alfabeto del Web semantico*, "Biblioteche oggi", 30 (2012), n. 3, p. 10, http://www.bibliotecheoggi.it/content/201200300701.pdf

Guerrini-Possemato 2015 — Guerrini, Mauro; Tiziana Possemato. *Creazione e ricerca dei dati nell'era digitale.* "Biblioteche oggi" 33(2015) n.4, p. 40

Heath – Bizer 2011 — Heath, Tom; Bizer, Christian. *Linked data evolving the Web into a global data space*. California, USA: Morgan & Claypool Publishers, 2011 http://linkeddatabook.com/editions/1.0/

Jenkins 2009 — Jenkins, Henry. *Confronting the challenges of participatory culture. Mediaeducation for the 21st century.* Cambridge, MA: MIT Press, 2009. https://mitpress.mit.edu/sites/default/files/titles/free_download/9780262513623_Confronting_the_Challenges.pdf

Malmsten 2012 — Malmsten, Martin. *National Library of Sweden* http://www.slideshare.net/geckomarma/ifla-2012-oclc-semantic-Web-roundtable

Wallis — Wallis, Richard. *OCLC* http://www.slideshare.net/rjw/library-linked-data-progress

Neil 2012 — Wilson, Neil. *British Library* http://www.slideshare.net/nw13/oclc-linked-data-roundtable-event-ifla-2012

W3C 2011 — W3C . *Library Linked Data Incubator Group Final Report*. 2011 http://www.w3.org/standards/semanticweb/ontology

Biblioteche digitali e Linked Open Data: contenuti, connessioni, link, argomentazioni

Maurizio Lana
Università del Piemonte orientale, Dipartimento di studi umanistici

Il Web è innervato da una trama di connessioni ipertestuali che ne costituiscono la principale modalità di organizzazione del contenuto e di funzionamento: si connette ciò che è simile; e cliccando sui link ci si sposta da un contenuto a un altro.
I Linked Open Data promettono un Web in cui esisterà un numero molto maggiore di connessioni oggettive, fattuali, tra i dati ma molto più interessanti e importanti sono le connessioni tra informazioni che (apparentemente) non hanno nulla in comune e che solo un interprete umano in forza di un'ipotesi o di un'argomentazione può mettere in relazione.

L'espressione Linked Open Data (LOD) descrive un particolare tipo di dati – Open Data, dati aperti – di cui si è iniziato a parlare alcuni anni fa soprattutto in relazione alla trasparenza politico-amministrativa delle istituzioni, orientata a permettere il controllo e la cooperazione da parte dei cittadini che, leggendo i dati per mezzo di strumenti software adeguati, possono capire come si comporta l'amministrazione ed eventualmente formulare proposte pertinenti ai problemi. Ma di Open Data si parla anche in relazione all'accesso aperto nel contesto della ricerca, perché pubblicare una ricerca e renderla accessibile senza però rendere accessibili i dati su cui essa si basa appare sempre più un modo limitato e riduttivo di intendere l'accesso aperto – come se interessasse solo la sua espressione conclusiva, la pubblicazione, e non anche il processo che a quell'espressione ha condotto. Tant'è vero che le licenze Creative Commons, che esprimono in termini legali formali le condizioni di accesso aperto a una risorsa, contemplano non solo le pubblicazioni in senso tradizionale e in senso ampio, ma anche i dati.
L'aggiunta dell'aggettivo *linked*, connessi, esplicita che i dati sono interconnessi: il contenuto di ogni raccolta di dati ha bisogno del contenuto di altre raccolte di dati per essere inteso in modo appropriato, compreso appieno nelle sue implicazioni, fruito in maniera ricca. Quasi si vorrebbe, ad analogia con il concetto di intertestualità, un concetto di "interdatalità": in quel caso connessioni tra i testi, in questo connessioni tra i dati. Nel mondo digitale le connessioni fino ad ora sono state pre-

SISTEMA MODUS

valentemente intese nel modo ipertestuale tipico: un frammento testuale intenzionalmente associato a un URL permette di raggiungere una risorsa. L'elemento rilevante è che in questo quadro di riferimento le connessioni avvengono *in genere* tra testi. Quando invece entrano in gioco i dati, i Linked Open Data, la testualità in certo modo "si ritira" in secondo piano perché in primo piano c'è l'insieme dei dati che sono frammenti da ricomporre, se si vuole ricavare informazione e costruire conoscenza[1].

La costruzione dei LOD è resa delicata da vincoli e limiti di tipo giuridico-commerciale perché spesso, anche se i metadati sono liberi (di un dato articolo ad accesso a pagamento è possibile copiare e condividere gli elementi descrittivi standard: autore, titolo dell'articolo, titolo della rivista, anno, numero, pagine, ecc.), i corrispondenti contenuti sono protetti (all'articolo menzionato chi non è abbonato alla rivista digitale non può accedere).

Link ipertestuali tradizionali

Dal punto di vista del discorso che si intende sviluppare, il Web contiene oggi due tipi di connessioni:

– connessioni costruite automaticamente: esse non implicano necessariamente LOD e collegano i dati per motivi fattuali, tali cioè da non essere oggetto di analisi, valutazione e interpretazione[2];
– connessioni costruite manualmente: che collegano dati oggetto di analisi, valutazione e interpretazione.

Sono soprattutto le connessioni del secondo tipo quelle che permettono la costruzione di ipertesti, cioè di scritti che sfruttano le possibilità argomentative *offerte dal mondo digitale* (senza dimenticare che possono esistere fuori del mondo digitale ipertesti che non richiedono necessariamente la disponibilità di un computer per essere navigati[3]). Per il lettore la disponibilità di computer e di ipertesti digitali ha a che fare solo con la velocità nel seguire i link e nel recepire i collegamenti, non con la loro costruzione concettuale, mentre dal punto di vista dell'autore la disponibilità di computer e risorse digitali facilita e sveltisce – quanto meno in ter-

[1] L'orizzonte in cui il discorso di questo articolo si colloca è quello delle scienze umane, non delle scienze MFN (matematiche fisiche naturali) o STEM (science technology engineering medicine), perché per le scienze umane la conoscenza è prevalentemente discorsiva mentre per le scienze MFN/STEM la conoscenza si può benissimo esprimere in tabelle di dati.

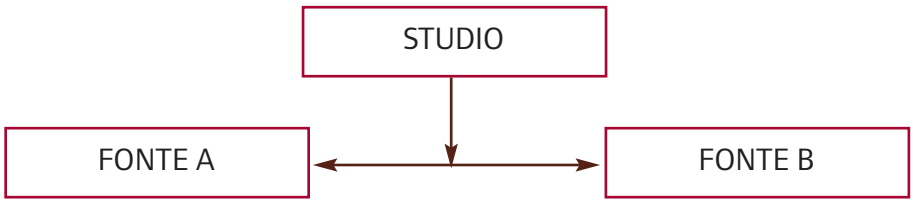
[2] Come accade per esempio nel sito di una libreria online.

[3] George Landow (*Hypertext: The convergence of contemporary critical theory and technology*, Johns Hopkins University Press, 1992) menziona a tale proposito la Bibbia e l'Enciclopedia per la trama dei rimandi interni; Ted Nelson (*The future of information*, ASCII, Tokyo, 1997) menziona i quotidiani, e i loro rimandi da pagine frontali a pagine interne; a questi tipi di scritti si possono senza dubbio aggiungere gli articoli e i saggi scientifici con il loro apparato di note.

SISTEMA MODUS

mini operativi – l'individuazione di connessioni e la creazione di link: collegando due differenti fonti si compie un atto interpretativo, per cui due entità distinte diventano parte di un'unità più ampia.

Per avere un esempio particolarmente incisivo di questa costruzione di conoscenza fondata su un'analisi e interpretazione di elementi disponibili nel mondo circostante si possono prendere in considerazione due eventi storici: l'uccisione di Ahmad Shāh Massoūd in Afghanistan il 9 settembre 2001 e l'attacco alle torri gemelle del World Trade Center l'11 settembre 2001. C'è chi sostiene che i due fatti sarebbero collegati da una comune matrice qaidista in quanto Massoūd era il rappresentante molto autorevole e militarmente efficace di un'opposizione antitalebana non integralista e dunque le due azioni avrebbero avuto il medesimo intento di colpire due nemici dell'integralismo islamico. Questa interpretazione non si può esprimere semplicemente collegando due pagine Web perché le fonti che testimoniano dei 2 attentati sono distinte e isolate e *il puro fatto di collegarle non spiega (non rivela) pienamente il motivo del collegamento*. Occorre un "punto intermedio" che riveli la connessione che l'autore-interprete individua negli avvenimenti. Se le fonti sono – come sono, in quanto fonti – non modificabili, questo punto intermedio ha in genere la forma di un terzo testo che istituisce esplicitamente la connessione:



Tra l'altro, banale a dirsi ma rilevante per il discorso che si sta sviluppando qui, le fonti vengono prima delle interpretazioni: vi sono quindi fonti che parlano dell'assassinio di Massoūd e fonti che parlano dell'attentato dell'11 settembre. Solo dopo gli eventi, riflettendo su di essi, qualcuno può leggerli individuandovi una connessione che un testo terzo successivamente esplicita verbalmente ed eventualmente per mezzo di uno o più link. Se gli elementi esteriori, fattuali, presenti nelle fonti e tali da connettere i due eventi sono a seconda dei punti di vista vaghi o molto frequenti o inesistenti:

- vaghi: questioni geopolitiche molto ampie e varie non riconducibili a elementi lessicali ricorrenti e chiaramente finalizzati, quale ad esempio lo scontro tra l'Islam radicale e i paesi occidentali;
- molto frequenti: i termini Islam e suicidio;
- inesistenti: nulla propriamente collega i due eventi, né la geografia, né i tempi, né gli attori;

l'intervento umano di analisi e interpretazione è in grado di individuare le connessioni laddove non sono esplicitamente visibili in primo piano o laddove le corrispondenze lessicali sono numerosissime ma solo in alcuni casi fortemente significative.

Per il caso menzionato, alcuni esempi di quello che abbiamo chiamato «terzo testo che istituisce esplicitamente la connessione» sono[4]:

- https://en.wikipedia.org/wiki/Ahmad_Shah_Massoud: He was assassinated, probably at the instigation *of al-Qaeda, in a suicide bombing* on September 9, 2001, just two days before the *September 11 attacks* in the United States.
- https://it.wikipedia.org/wiki/Ahmad_Shah_Massoud: Massoud fu assassinato in un attentato suicida il 9 settembre 2001 a Khvājeh Bahā' od-Dīn da due tunisini che si fingevano giornalisti di una emittente marocchina. I due terroristi erano stati reclutati, secondo una teoria, a Bruxelles dall'emiro loro connazionale Saif Allah Ben Hassine, capo e fondatore dell'organizzazione salafita Ansar Al Sharia. *Altre teorie accreditano una matrice qaidista dell'attentato*.
- https://it.wikipedia.org/wiki/Fronte_islamico_unito_per_la_salvezza_dell%27Afghanistan: Il 13 settembre 2001 venne confermata la morte di Massoud, ucciso quattro giorni prima da infiltrati di al-Qa'ida che si erano presentati all'organizzazione come giornalisti.
- https://en.wikipedia.org/wiki/Northern_Alliance: On September 9, 2001, two Arab *suicide attackers*, allegedly belonging to Al Qaeda, posing as journalists, detonated a bomb hidden in a video camera while interviewing *Ahmed Shah Massoud* in the Takhar province of Afghanistan. Commander Massoud died in a helicopter that was taking him to a hospital. He was buried in his home village of *Bazarak* in the *Panjshir Valley*. The funeral, although taking place in a rather rural area, was attended by hundreds of thousands of mourning people. The assassination of Massoud is considered to have a strong connection to the *attacks in the U.S.* two days later.
- http://www.thenational.ae/news/world/south-asia/ahmad-shah-massoud-assassinated-by-al-qaeda-but-no-friend-of-the-us (6/9/2011): *Osama bin Laden is widely believed to have orchestrated Massoud's assassination* as a favour to the Taliban, whose protection would be critical after the 9/11 hijackers, already moving into place in Boston, Newark and suburban Washington, carried out their attacks.
- http://video.sky.it/news/mondo/afghanistan_massud_prima_vittima_di_al_qae-

[4] La ricerca è stata effettuata con Google Search il 25 settembre 2015, utilizzando soltanto le parole massud e qaeda; sono stati utilizzati i primi 8 esiti. I nomi in gioco presentano numerose varianti: massud o massoud o massoūd, qaeda o qaida o qa'ida. Non sono state esplorate tutte le combinazioni fra queste varianti e altre ancora in quanto non rilevanti per questo discorso, in cui si discute su come si costruiscono le connessioni di senso e come esse si esprimono con gli strumenti del mondo digitale.
La data riportata di alcune fonti accanto all'URL è quella della pagina Web.

da/v96788.vid (9/9/2011): *Ucciso il 9 settembre 2001 da un attacco suicida di Al Qaeda*, in Afghanistan viene ancora oggi ricordato e celebrato Ahmad Shah Massud, oppositore del regime talebano.

- http://it.euronews.com/2011/09/09/afghanistan-ricorda-massoud-nemico-di-al-qaeda-ucciso-10-anni-fa/ (9/9/2011): Leone del Panjshir, è da questo bastione imprendibile e sua terra natale che Massoud resiste prima ai sovietici e poi ai taleban fino all'*omicidio, commesso in un attentato suicida da due falsi giornalisti a ridosso dell'11 settembre*.
- http://www.aljazeera.com/indepth/opinion/2014/09/afghanistan-shadow-ahmad-shah-mas-2014997826874331.html (9/9/2014): The story of 9/11 is in reality the story of the legendary Afghan freedom fighter Ahmad Shah Massoud. […]Osama bin Laden is widely believed to have orchestrated Massoud's assassination as a favour to the Taliban, whose protection would be critical after the 9/11 hijackers, already moving into place in Boston, Newark and suburban Washington, carried out their attacks.

Le espressioni usate per indicare la connessione tra gli eventi sono molto varie:

- altre teorie accreditano una matrice qaidista dell'attentato;
- ucciso il 9 settembre 2001 da un attacco suicida di Al Qaeda;
- Osama bin Laden is widely believed to have orchestrated Massoud's assassination;
- just two days before the September 11 attacks in the United States;
- omicidio, commesso in un attentato suicida da due falsi giornalisti a ridosso dell'11 settembre;
- the assassination of Massoud is considered to have a strong connection to the attacks in the U.S. two days later;
- Osama bin Laden is widely believed to have orchestrated Massoud's assassination as a favour to the Taliban, whose protection would be critical after the 9/11 hijackers carried out their attacks.

Si può facilmente osservare che queste espressioni esprimono la connessione, ma solo qualche autore la traduce in link (evidenziati in corsivo), come si può vedere nei passi estesi riportati più sopra[5].

Il link, quando presente, è utilizzato nel modo onnipresente che si potrebbe definire "confermativo": un'affermazione viene linkata a una fonte che la conferma. La conferma può configurarsi come dettaglio (il link porta a dettagli informativi su ciò che l'affermazione di partenza esprime in modo conciso) o come approfondimento

[5] Non sono stati riportati i link fattuali e ovvi, qui segnalati in corsivo, presenti in alcune pagine Wikipedia, come ad esempio: <https://it.wikipedia.org/wiki/Ahmad_Shah_Massoud>: «Massoud fu assassinato in un attentato suicida il 9 settembre *2001* a Khvājeh Bahāʾ od-Dīn da due *tunisini* che si fingevano giornalisti di una emittente *marocchina*. I due terroristi erano stati reclutati, secondo una teoria, a *Bruxelles* dall'*emiro* loro connazionale Saif Allah Ben Hassine, capo e fondatore dell'organizzazione *salafita* Ansar Al Sharia. Altre teorie accreditano una matrice *qaidista* dell'attentato.» In tutti questi casi i link rimangono all'interno di Wikipedia – sono

(il link porta ad un'esposizione criticamente più approfondita della formulazione offerta dall'affermazione di partenza). Questo modo d'uso del link apparentemente molto semplice si presta ad usi complessi, perché si potrebbe avere un testo che afferma: «Non è sostenibile l'affermazione che l'assassinio di Massud sia connesso con l'attentato dell'11 settembre alle Torri gemelle (<http://www.nytimes.com/learning/general/onthisday/big/0911.html>»: in questo caso il link ha il consueto uso confermativo (porta a una pagina Web del NYT in cui viene descritta in dettaglio l'affermazione concisa «attentato dell'11 settembre …»): ma si trova all'interno di un'argomentazione in negativo («non è vero che …»). Il link (il meccanismo del link), non dice nulla sul tipo di argomentazione di cui fa parte – con il risultato che non è possibile formulare ricerche complesse in cui le pagine Web che contengono una determinata stringa di testo come *anchor text* (anche noto come *hot text*, testo cliccabile)[6] vengano ulteriormente selezionate in base al fatto che il link abbia un significato (una funzione) o un'altra. Eppure nella concezione originaria del Web di Berners-Lee nel 1989[7] il Web era «a large hypertext database with typed links»[8] (ma i typed links nello sviluppo successivo si persero).

Link = connessione?

Uno dei motivi d'interesse per i Linked Open Data sta nel fatto che l'attenzione si amplia dalla componente *open* – i dati sono liberamente accessibili – che compare per prima, alla componente *linked* – i dati sono interconnessi – perché dati isolati servono a poco/dicono poco. I dati *open* che espongono la spesa sanitaria di una Regione sono poco significativi se non si accompagnano a dati che descrivono il bilancio di quella Regione; per entrambi i tipi di dati le condizioni di comprensibilità (di interpretabilità) migliorano se se ne possiede una serie storica, e migliorano ulteriormente se è possibile confrontarli con altri dati aperti del medesimo tipo di altre Regioni. Fin qui la connessione tra i dati è una conoscenza pregressa posseduta a priori da coloro che di questi dati si interessano.

dei «Wikilink» – e non portano a fonti storiche sugli eventi. E anzi nelle indicazioni per gli autori Wikipedia raccomanda, <https://it.wikipedia.org/wiki/Aiuto:Tutorial/Link>: «Se vuoi introdurre link a siti esterni ai progetti Wikimedia, ti raccomandiamo di farlo in fondo alla voce, dove troverai (o potrai scrivere tu) una sezione chiamata collegamenti esterni».

6 È l'operatore *allinanchor*: una query formulata con *allinanchor: massoud* cercherebbe tutte le pagine Web in cui la parola massud è parte di un testo cliccabile.

7 Tim Berners-Lee, *Information management. A proposal*, <http://www.w3.org/History/1989/proposal.html>.

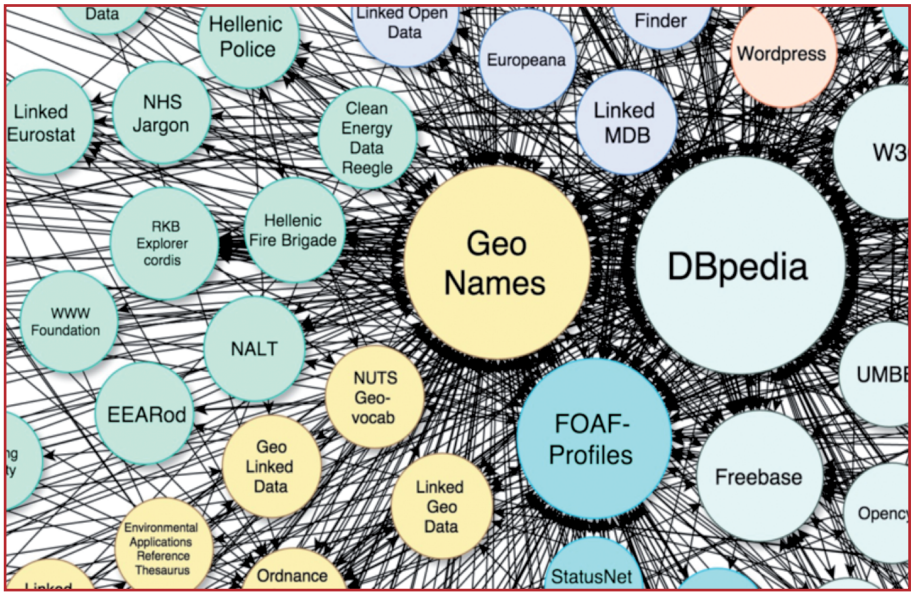
8 E in un documento storico senza data che si trova nel sito wed Web Consortium, <http://www.w3.org/DesignIssues/Topology.html> si legge: «Should links be typed? A typed link carries some semantic information, which allows the system to manage data more efficiently on behalf of the user. A default type («untyped») normally exists in some form when types are implemented. See also a list of some types. (Should a link be allowed to have many types? (-JFG) I don't think so: that should be represented by more than one link.(- TBL). Link typing helps with the generation of graphical overviews, and with automatic tracing .» TBL è Tim Berners-Lee.

SISTEMA MODUS

Nel caso dei LOD generalmente il link non costituisce una connessione in senso stretto, cioè l'esito di un atto interpretativo, ma solo l'ampliamento di una rete di riferimenti fattuali. Si può partire da un esempio semplice, tratto dal LOD Hackaton 2015 @ NEXA[9]: dati.piemonte.it contiene open data sui musei del Piemonte, tra cui un campo «orari di apertura» costituito da informazioni semi-strutturate. Passando i dati ad un'apposita applicazione si possono separare in colonne a parte le chiusure festive, le aperture straordinarie e serali, il tipo di ingresso (a pagamento/gratuito) che originariamente sono informazioni di tipo testuale descrittivo fra di loro disparate, creando una connessione tra musei in base al tipo di eventuale apertura festiva, oppure in base al tipo di apertura serale, e così via. Si può in certo modo dire che i dati sono stati linkati con un procedimento tanto indiscutibile quanto ovvio: indiscutibile perché il link connette i musei che hanno la medesima eventuale apertura serale e il dato che descrive tale apertura esiste, ma non è facile connetterlo con altri perché deve essere riconosciuto ed estratto dal contesto testuale in cui è espresso; ovvio perché il procedimento esplicita e mette in ordine ciò che già si sapeva in modo disordinato e quindi non costruisce un'interpretazione esposta a discussione, smentita, critica.

Un esempio più complesso è offerto dalla nota mappa della LOD cloud[10] di cui qui viene presentato un frammento.

I cerchi rappresentano i dataset i cui dati sono connessi con quelli di altri dataset. Come si può osservare, i dataset reciprocamente interconnessi sono dei tipi più diversi e hanno al centro (come dataset che ricevono e danno il maggior numero di



[9] http://nbviewer.ipython.org/gist/steko/0d87f5b9e32691de28be.

[10] Linking Open Data cloud diagram 2014, by Max Schmachtenberg, Christian Bizer, Anja Jentzsch and Richard Cyganiak, <http://lod-cloud.net/>.

link) Geonames e DBpedia, cioè due risorse informative fattuali: ovunque vi sia un nome di luogo esso può essere linkato a Geonames, che ne darà la descrizione della struttura geografica (regione, stato, continente ecc.) dentro cui quel nome si colloca; o a DBpedia che costituisce il punto di entrata nei contenuti di Wikipedia organizzati sotto forma di LOD. Viene dunque creata una fitta rete di link senza che vi siano atti interpretativi, senza cioè un'intenzionale individuazione di connessioni altamente significative: il fatto che il nome «Catania» presente in una notizia di un quotidiano online che parla di prospettive di sviluppo di STMicroelectronics, azienda di elettronica di Catania, sia linkato a Geonames e di lì ad un articolo scientifico in cui si parla della cattedrale di S. Agata a Catania, non significa che vi sia una connessione tra lo sviluppo dell'azienda e la cattedrale della città. Ma se in quello stesso articolo si dicesse che a causa di difficoltà economiche l'azienda ridurrà le sue azioni di mecenatismo verso la città in cui ha sede, e questo comportasse un rallentamento dei lavori di restauro della cattedrale della città, allora i link citati sopra si presenterebbero alla mente del lettore/esploratore/navigatore come connessioni fortemente significative.

Le connessioni arricchiscono di senso il testo che le contiene

Nel mondo LOD l'esistenza di *n* link non corrisponde ad altrettante interpretazioni espresse da menti umane che hanno investigato i dati, perché la creazione del link, come si è detto e come è noto, avviene automaticamente. È altresì vero che l'atto interpretativo consistente nell'individuazione di una connessione tra dati, tra unità di informazione, si esprime nella creazione di uno o più link non solo nel mondo digitale ma anche nel mondo fisico della carta; ma se le interpretazioni connettive si esprimono operativamente nei link, nel Web dei LOD non ogni link esprime un'interpretazione.

Peraltro, sulla scorta dell'esperienza in corso nel progetto Geolat – geografia per la letteratura latina[11]- questa riflessione può essere arricchita di approfondimento critico. Geolat annota ontologicamente i testi latini individuando in essi i nomi geografici e pubblica come LOD la relazione tra nome geografico e estremi del passo annotato[12]. Da tempo esiste online Pleiades[13], un gazzettiere geografico del mondo antico basato sul Barrington Atlas[14] che è la fonte autoritativa sulla geografia del mondo antico. Pleiades pubblica come LOD tutte le informazioni che contiene, tra cui le coordinate GPS delle località. Tra Geolat e Pleiades potrà avvenire un'inter-

[11] http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/lana/

[12] I LOD pubblicati sono anche altri ma non sono così efficaci per lo sviluppo di quest'argomentazione.

[13] http://pleiades.stoa.org.

[14] *Barrington atlas of the Greek and Roman world. Map-by-map directory*, edited by Richard J. A. Talbert, 2000, Princeton.

connessione di questo tipo: Geolat si connetterà ai LOD di Pleiades e di essi importerà le informazioni GPS, mentre Pleiades potrà connettersi ai LOD offerti da Geolat per importare la relazione tra nomi geografici e citazioni dei passi che li contengono. In altri termini le "transazioni LOD" nel caso di Geolat e Pleiades non daranno luogo ad una rete di link automatici della massima estensione possibile ma ad una selezione mirata e funzionale ai reciproci scopi:

- Geolat riceverà da Pleiades un dato geografico di fondamentale importanza evitando di scriverlo in proprio esponendosi ad errori materiali di battitura;
- Pleiades riceverà da Geolat un dato fondamentale per individuare dove effettivamente ricorrono nei testi i nomi lì elencati;
- in Geolat l'individuazione del dato (il nome geografico) non è frutto di una procedura automatica di NER (Named Entity Recognition) bensì di un'attività di analisi e annotazione svolta da studiosi[15].

Lo scopo di questo modo di procedere, rispetto a quello prevalente che si potrebbe definire *computer-based*, è di porre in primo piano l'attività valutativa e interpretativa dello studioso e quindi di produrre una risorsa efficace su cui altri studiosi possano successivamente contare: efficace perché, espresse sotto forma di LOD, contiene solo connessioni significative.

L'annotazione arricchisce perché aumenta la connettibilità

Nella prospettiva qui delineata si può ricordare che GO!, l'ontologia utilizzata in Geolat per l'annotazione dei nomi geografici[16], permette tra l'altro di descrivere gli eventi storici che si sono verificati in uno specifico luogo dettagliando, quando siano noti, i nomi dei personaggi in esso coinvolti[17]. Ciò permette di ipotizzare e progettare l'interconnessione con annotazioni che utilizzino PRoles[18], un'ontologia che descrive i ruoli politici – il che potrebbe apparire talmente ovvio da risultare banale. Ma come si è detto, quando in GO! l'annotazione del nome di un determi-

[15] Lo studioso in realtà utilizza anche gli esiti di un parsing NER-oriented effettuato sui testi, ma semplicemente allo scopo di uno spoglio iniziale che aiuti a velocizzare le operazioni ripetitive, quale ad esempio annotare il nome "Roma"; mentre di fronte a nomi come "Alessandria" occorre disambiguare; e di fronte a stringhe come "uxellodunum" probabilmente difficile da riconoscere automaticamente come nome di luogo occorre il riconoscimento fatto dallo studioso.

[16] Alla data di pubblicazione di questo articolo l'ontologia si trova online all'URL: <https://www.dropbox.com/sh/j2uaq60alocll97/AAAWfcyybPAL7uNJLaY-SOxea?dl=0> in quanto il servizio PURL non accetta la creazione di nuovi domini. Autori C. Corcione, P. De Caro, S. Naro, con collaborazione di D. Magro, M. Lana, T. Tambassi.

[17] Non ci si stupisca di questa apparentemente strana commistione di conoscenza geografica e dati storici, perché uno dei motivi per cui un punto dello spazio acquista la rilevanza specifica che si esprime nell'attribuzione di un nome è l'attività umana di significazione spesso connessa agli eventi che vi si svolgono.

[18] <http://www.essepuntato.it/2013/10/politicalroles>, di M. Daquino, S. Peroni, e F. Tomasi.

SISTEMA MODUS

nato luogo contiene il nome di un personaggio storico protagonista di un evento svoltosi in quel luogo (ad esempio il nome geografico è Rubicone, l'evento è l'attraversamento del fiume da parte di Cesare con il suo esercito), *il nome di quel personaggio non è presente nel testo ma viene recuperato grazie ad un lavoro di approfondimento di secondo livello da parte dello studioso* (primo livello: dal nome di luogo all'evento storico che vi si è svolto; e dall'evento al nome del personaggio storico coinvolto o protagonista). Ciò è un chiaro esempio del valore della capacità analitico-interpretativa dello studioso, più sottile e sicura di un procedimento *computer-based*; ma anche dell'importanza di annotare il testo per mezzo di ontologie, che permettono di descrivere il testo in modo ricco e ramificato, "sinaptico" nel senso della capacità di esplicitare e creare connessioni: perché senza ontologia il nome di Cesare non potrebbe essere correttamente messo in gioco dove ricorresse solo il nome Rubicone, perdendo così l'opportunità di creare connessioni di senso verso altri testi che contengono annotazione ontologica di personaggi con ruoli politici.

Conclusioni

Il modo oggi diffuso di intendere i LOD evidenzia una caratteristica tipica del mondo digitale e ben nota fin dalle origini: produrre dati (Linked Data) è molto facile ma la ricerca non consiste nel produrre dati con procedure automatiche sostanzialmente incontrollate, bensì nel produrre dati che rispondano ad un'ipotesi interpretativa, per poi connetterli e pubblicarli (Linked Open Data) affinché l'ipotesi, venendo comunicata, possa essere condivisa e verificata.

> *Hypertextual connections innervate the Web, and at the same time they constitute the main way the content is working: what is similar is linked and by clicking on a link one goes from one content to another.*
> *The Linked Open Data promise a Web where a much bigger number of factual, objective connections can be set up among data. Furthermore, connections among pieces of information apparently having nothing in common, which only human interpretation can relate the one to the other thanks to a hypothesis or an argumentation, prove to be by far more interesting and relevant.*

L'ultima consultazione dei siti Web è avvenuta nel mese di dicembre 2015.

SISTEMA MODUS

Web dei dati alla Biblioteca nazionale centrale di Firenze[1]

Anna Lucarelli
Biblioteca nazionale centrale di Firenze

Le Biblioteche nazionali di tutto il mondo stanno lavorando per favorire l'accesso alle proprie risorse da parte degli utenti del Web, al di là dei confini ristretti dei cataloghi. L'obiettivo è mettere a disposizione non solo opere ed edizioni, ma anche dati di qualità strutturati e attribuiti in fase di catalogazione o di digitalizzazione, rendendoli riusabili e disponibili nei formati del Web semantico. I linked open data sono una nuova e straordinaria opportunità e potranno dare un significato nuovo al ruolo sociale e pubblico che le Nazionali rivestono. La Biblioteca nazionale centrale di Firenze ha inaugurato le sue esperienze di open data con il Thesaurus del Nuovo soggettario, strumento evoluto nella direzione del multilinguismo e interoperabile con data set di altre amministrazioni pubbliche e istituzioni della memoria.

Open data: esperienze di biblioteche nazionali

In ogni luogo del mondo le biblioteche nazionali sono istituzioni che conservano, tutelano e diffondono la memoria culturale e il patrimonio documentario del proprio Paese. Hanno il compito non solo di ospitare la produzione editoriale nazionale, ma anche di coordinarne l'informazione, giocando un ruolo importante anche a livello internazionale con il mantenimento di bibliografie nazionali, di authority files ecc.

Le loro funzioni si sono da un lato sempre più precisate, dall'altro si sono allargate per soddisfare, con prodotti nuovi, un'utenza più ampia e diversa rispetto a quella tradizionale, l'utenza del Web. E sappiamo che gli utenti del Web hanno nuove aspettative e nuove abitudini.

[1] Relazione presentata alla Tavola Rotonda *Biblioteche digitali verso il futuro: accesso ai linked open data-Book to the future* (coordinata da Anna Maria Tammaro).

SISTEMA MODUS

Le biblioteche nazionali, grazie alla normativa sul deposito legale, sono a contatto con l'intero panorama editoriale di un Paese - libri, periodici, opere grafiche e altre risorse documentarie - e dunque si muovono in un contesto molto vicino a quello rappresentato dal Salone del libro di Torino del 2015, nell'ambito del quale si è svolta la Tavola rotonda su "Biblioteche digitali verso il futuro".

Sappiamo che l'universo digitale ha coinvolto le biblioteche per gradi. In principio, i formati elettronici hanno riguardato le schede dei cataloghi, le liste di autorità, alcuni strumenti; poi, nello spazio virtuale e fisico delle biblioteche, hanno iniziato a evolvere funzioni di sviluppo di alcuni servizi, si sono attivate procedure legate all'acquisizione e conservazione di documenti digitalizzati o nati in formato digitale, iniziative che hanno avuto una diffusione massiccia proprio negli ultimi anni anche grazie alle collaborazioni avviate con progetti ben noti come Google Books al quale, in Italia, hanno aderito le Biblioteche nazionali centrali di Roma (coordinatrice del progetto) e di Firenze, la Biblioteca nazionale Vittorio Emanuele III di Napoli e recentemente altre biblioteche[2].

A livello internazionale, ci si è chiesto come fornire maggior accesso e visibilità ai dati prodotti dalle biblioteche, al di là dei confini ristretti dei cataloghi. Il passo avanti più significativo è stato quello di comprendere che ciò che le biblioteche possono mettere a disposizione non sono soltanto le opere e i testi che le veicolano, ma anche i dati che le riguardano, recependo così appieno il passaggio di prospettiva, per certi aspetti rivoluzionario, del Web semantico, il "Web dei dati", come lo ha appunto notoriamente definito Tim Berners-Lee.

La diffusione dei metadati, intesi come dati strutturati che descrivono altri dati e consentono ai computer, tramite particolari applicazioni, di riconoscere relazioni tra entità, favorendo il reperimento di informazioni, si inserisce appieno nella logica degli open data, dati aperti collegati fra loro (linked open data), grazie a licenze che ne garantiscono accessibilità e riuso[3]. E d'altra parte, la consapevolezza che i linked data siano una nuova opportunità per istituzioni e utenti è sempre più condivisa ed è alla base dell'allestimento di aggregatori di risorse digitali e di ben noti progetti internazionali come Europeana[4].

Così le biblioteche nazionali di tutto il mondo si stanno facendo, in varia misura,

[2] Il sito della BNCR dedicato a Google Books è: <http://www.bncrm.librari.beniculturali.it/index.php?it/832/progetto-googlebooks>; notizie più recenti sul progetto si trovano anche nell'intervista rilasciata, a fine settembre 2015, da Andrea De Pasquale, direttore della BNCR:<http://fontegaia.hypotheses.org/1230>.

[3] Sul tema degli open data esiste una ricchissima bibliografia. Si segnala: Mirna Willer – Gordon Dunsire, *Bibliographic information organization in the semantic Web,* Oxford [etc.]: Chandos Publishing, 2013; fra i contributi italiani più recenti: Mauro Guerrini –Tiziana Possemato, *Linked data per biblioteche, archivi e musei: perché l'informazione sia del Web e non solo nel Web,* con un saggio di Carlo Bianchini e la consulenza di Rosa Maiello e Valdo Pasqui, prefazione di Roberto delle Donne, Milano: Bibliografica, 2015.

[4] http://www.europeana.eu/portal/.

SISTEMA MODUS

sempre più promotrici di open data. Si sta iniziando anche in Italia a lavorare affinché i metadati da esse prodotti siano più visibili, più facili da ricercare, linkabili e riusabili direttamente. Se da un lato le biblioteche non sono più gli unici produttori di dati bibliografici (come dimostrano le esperienze che, ad esempio, nell'ambito di Wikidata, si stanno sperimentando grazie a Wikibase)[5], d'altro lato l'enorme massa di metadati, generalmente e tradizionalmente di qualità, che esse producono, se convertiti e messi a disposizione nei linguaggi del Web semantico, se resi interoperabili e aperti, potranno avere un'applicabilità ampia e diventare utili a un numero di cittadini ben superiore a quello attuale. Questo scenario offre alle biblioteche l'opportunità di svolgere nuove modalità di servizio pubblico (ad esempio, condividendo dati con altri *data set* della pubblica amministrazione) e di dare significato agli investimenti che impiegano nella catalogazione e nella digitalizzazione, evitando costi di duplicazione e generando forme di risparmio. D'altra parte, sul valore degli open data come servizio per i cittadini, ci sono ormai esperienze importanti come, ad esempio, quelle realizzate dalla Regione Piemonte, dalla Regione Emilia Romagna o dal Comune di Firenze.

Dunque le attività catalografiche svolte da sempre nell'ambito delle biblioteche sono tutt'altro che divenute inutili, ma gli investimenti impiegati avranno maggior valore se i dati relativi alle risorse descritte, oltre che affidabili, si baseranno su modelli concettuali condivisibili (come FRBR), adotteranno standard del *semantic Web*, ad esempio, rendendosi disponibili nelle triple del linguaggio RDF (*Resource Description Framework*)[6].

Come già accennato, sono numerose le esperienze e le sperimentazioni di *linked data service* avviate da biblioteche nazionali di tutto il mondo: fra i casi prototipali ed esemplari si citano solitamente quelli di Svezia, Germania, Francia, Regno Unito, USA ecc.[7]. E vanno senz'altro guardate con grande interesse le recenti iniziative dell'Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche (ICCU) che ha costituito un gruppo di lavoro per valutare la possibilità di strutturare set di dati prodotti nell'ambito di SBN come linked open data[8].

[5] Sull'argomento, Giovanni Bergamin, *La sfida di Wikidata alle biblioteche*, Convegno nazionale *Sfide e alleanze fra biblioteche e Wikipedia*, Biblioteca nazionale centrale di Firenze, 28 novembre 2014, <http://www.slideshare.net/GiovanniBergamin/la-sfida-di-wikidata-alla-biblioteche>.

[6] Il tema dell'apporto che le biblioteche, specialmente Nazionali, possono dare a livello sociale tramite i propri dati aperti è stato già affrontato da bibliotecari ed esperti di vari paesi. Un riferimento a tali contributi si trova anche in Anna Lucarelli, *Nuove scommesse della BNCF: Wikipediani in residence, Wikisource e altro ancora,* «Dig*Italia*», 9 (2014), n. 2, p. 100-106. <http://digitalia.sbn.it/article/view/1292/849>. Sul valore sociale dei linked data, anche Francesca Di Donato, *Lo Stato trasparente. Linked open data e cittadinanza attiva*, Pisa : ETS, 2010.

[7] Interessante, per quanto riguarda la BNF, la rubrica aggiornata *Web de données, Web sémantique*, <http://www.bnf.fr/fr/professionnels/innov_num_web_donnees.html>.

[8] < http://www.iccu.sbn.it/opencms/opencms/it/main/attivita/gruppilav_commissioni/pagina_0005.html>. Sul tema, si segnala anche l'intervento di Patrizia Martini nella lista AIB-CUR in data 30 settembre 2015.

SISTEMA MODUS

La Biblioteca nazionale centrale di Firenze (BNCF) ha inaugurato le sue esperienze di open data grazie a uno strumento che allestisce per l'intero sistema delle biblioteche italiane: il Thesaurus del *Nuovo soggettario*.

Nuovo soggettario e open data: da strumento catalografico a fornitore di dati aperti

Come la Library of Congress, la Bibliothèque nationale de France e molte altre grandi biblioteche, anche la BNCF cura il principale sistema nazionale impiegato nell'indicizzazione per soggetto, la cui componente centrale è un Thesaurus multidisciplinare online di oltre 56.000 termini usabili nella catalogazione di risorse varie e dunque in un'ottica fortemente orientata alla condivisione di linguaggi con musei, archivi, fototeche ecc.[9].

Lo strumento è disponibile online dal 2007 e dal 2013 con un'interfaccia anche in inglese[10].

È interessante raccontare di come uno strumento nato come un "attrezzo" del mestiere dei bibliotecari, dunque con fini esclusivamente catalografici e inserito pienamente in una prassi bibliotecaria di stampo tradizionale, possa in realtà evolvere con caratteristiche impreviste fino a qualche anno fa e con funzionalità che stimolano il collegamento non solo fra biblioteche e altre istituzioni della memoria, ma anche con altre amministrazioni pubbliche che stanno sviluppando progetti analoghi nella stessa direzione.

Il Thesaurus è integrato con il catalogo della BNCF ed è integrabile con qualsiasi Opac delle altre biblioteche che lo adottano, consentendo in questo modo navigabilità e collegamenti diretti con le notizie bibliografiche. Si alimenta accogliendo terminologia usabile nell'indicizzazione e proposta dalla rete di biblioteche, istituzioni e centri di ricerca che partecipano al progetto.

Rappresenta dunque un esempio di come si possa mettere a disposizione di biblioteche digitali, anche specialistiche, *tools* prodotti da una biblioteca nazionale e di come strumenti di accesso all'informazione si possano sviluppare sfruttando competenze e risorse professionali di varia natura, grazie alla cooperazione fra istituzioni diverse, generaliste e specialistiche, pubbliche e private[11].

Come è noto, i termini del Thesaurus sono organizzati secondo le categorie e le relazioni previste dagli standard internazionali (è in corso l'adeguamento all'ultimo standard ISO 25964 che ha sostituito i precedenti 2788 e 5964), sono corredati da

[9] Della possibile integrazione di metadati prodotti in contesti documentari diversi si sta occupando un gruppo di lavoro formatosi nell'ambito del coordinamento del MAB Toscana, < http://www.mab-italia.org/index.php/comitatati/mab-toscana>. Il lavoro del Gruppo è confluito nel report: *Verso l'integrazione tra archivi, biblioteche e musei. Alcune riflessioni*, a cura del MAB Toscana – Gruppo Linguaggi (in via di pubblicazione).

[10] Per la descrizione del sistema di indicizzazione: <http://thes.bncf.firenze.sbn.it/index.html>; per l'accesso al Thesaurus: <http://thes.bncf.firenze.sbn.it/ricerca.php>.

[11] Per l'elenco dei partner del progetto: <http://thes.bncf.firenze.sbn.it/enti.htm>.

SISTEMA MODUS

note, da collegamenti con sinonimi e forme non più preferite (varianti storiche), dall'indicazione di corrispettivi numeri della *Classificazione decimale Dewey*, nella versione aggiornata della *WebDewey italiana*[12], e inoltre dalla segnalazione dei repertori, enciclopedie, database, usati come fonti per il controllo di morfologie e significati (Fig. 1).

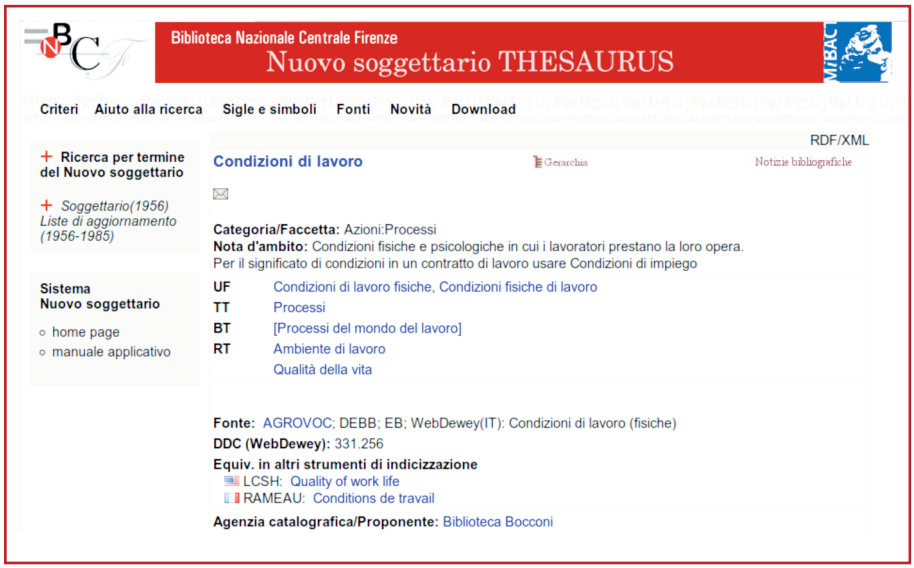


Figura 1: *Esempio di un termine del Thesaurus del Nuovo soggettario*

Dati aperti e interoperabilità del *Nuovo soggettario*

Come già accennato, il Thesaurus negli ultimi anni si è sempre più evoluto nella direzione, oltre che del multilinguismo, anche dell'interoperabilità con altri strumenti e sistemi di organizzazione della conoscenza. Colloquia, attraverso link, con gli equivalenti in inglese di *Library of Congress Subject Headings* – LCSH (oltre 10.000 collegamenti reciproci), con migliaia di equivalenti in francese previsti dal sistema *Rameau* della Bibliothèque nationale de France (quasi 8.000), con circa 13.000 lemmi della versione in italiano di *Wikipedia*, con thesauri online (AAT, *Agrovoc*, *Eurovoc*, *Polimoda*) e con altre risorse digitali come l'*Enciclopedia Italiana* Treccani, sistemi di classificazione (BIAnet, DoGI). I dati numerici indicati, aggiornati al mese di dicembre 2015, sono tutti in costante accrescimento. Nel corso del 2016 sarà implementata anche la navigabilità con *WebDewey italiana*, le cui notazioni compaiono già nei record di circa 5.000 termini del Thesaurus (Fig. 2).

[12] http://www.aib.it/pubblicazioni/webdewey-italiana/.

SISTEMA MODUS

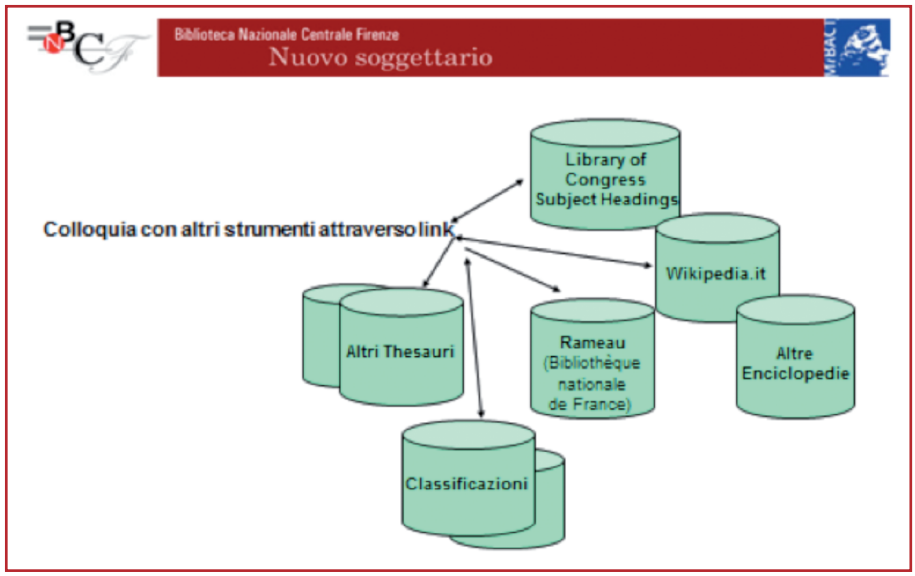


Figura 2: *I collegamenti del Nuovo soggettario con altri strumenti e data set*

La maggior parte dei collegamenti attivati dal *Nuovo soggettario* sono possibili perché i metadati del Thesaurus dal 2010 sono rilasciati nel particolare formato SKOS, sviluppato dal W3C (Semantic Web Deployment Working Group - SWDWG) e basato sul linguaggio RDF[13]. Come è noto, si tratta di uno standard molto efficace per far colloquiare sul Web i thesauri fra di loro o con altre strutture concettuali, come tassonomie o classificazioni. SKOS, infatti, può semplificare la comunicazione e collegare mondi che altrimenti rimarrebbero separati, grazie al fatto che la sua struttura si basa su concetti, indipendentemente dalla loro espressione linguistico-letterale e che tali concetti e le relazioni fra loro sono espressi con triple RDF. Il concetto viene associato (tramite il predicato *rdf:type)* alla classe specifica *skos:Concept* ed è identificato in modo univoco con un *Uniform Resource Identifier*. Così, l'assegnazione di un valore univoco ai concetti facilita l'interoperabilità tra *Knowledge Organization Systems* (KOS) differenti, cioè la possibilità di mappare entità semantiche di schemi concettuali diversi[14].
La conversione in SKOS del *Nuovo soggettario* ha seguito varie fasi, non è stata priva di problemi, alcuni tuttora aperti, ma il lavoro è comunque evoluto e sta facendo tesoro degli ultimi allineamenti stabiliti fra SKOS e il già citato nuovo standard ISO 25964[15].

[13] http://www.w3.org/2004/02/skos/.
[14] Sull'argomento già Giovanni Bergamin – Anna Lucarelli, *The Nuovo soggettario as a service for the linked data world*, «JLIS.it», 4 (2013), n. 1, <http://leo.cineca.it/index.php/jlis/article/view/5474/7903>.
[15] Sulle corrispondenze fra i due standard: <http://www.niso.org/apps/group_public/download.php/12351/Correspondence%20ISO25964-SKOSXL-MADS-2013-12-11.pdf>.

Il Thesaurus del *Nuovo soggettario*, insomma, punta a soddisfare quelli che vengono usualmente definiti come i requisiti di base dei dati aperti. È indicizzabile dai motori di ricerca, è disponibile in un formato aperto e standardizzato che facilita la sua consultazione, è rilasciato attraverso licenze libere che ne incentivano la diffusione e il riuso da parte di soggetti interessati[16].
Oltre ai fronti già citati, se ne dischiudono altri come, ad esempio, quello dell'impiego del Thesaurus come strumento di indicizzazione automatica di risorse digitali, versante su cui si sono già avviate sperimentazioni[17].

Implementazione dei linked data nel *Nuovo soggettario*

Le corrispondenze del *Nuovo soggettario* possono essere attivate mediante l'indicazione di altri strumenti nel campo "Fonte" previsto nel record di ogni termine. Se lo strumento citato è disponibile in SKOS, le relazioni si arricchiscono tramite *skos:closeMatch*, se invece non è disponibile in questo formato, l'informazione si perde, pur restando citata nel campo "Fonte" la sigla relativa allo strumento linkabile. Il collegamento può avvenire anche tramite l'indicazione di equivalenza esplicitata in altri campi, come avviene per i corrispettivi previsti da LCSH, con l'attivazione di *deep link* allo strumento citato.
È interessante rilevare che la collaborazione con la maggior parte delle istituzioni che producono dati collegati è precedente alla pubblicazione dei dati in formato SKOS, a riprova del fatto che il piano delle *policies* resta comunque fondamentale anche nel mondo degli open data.
E comunque non tutte queste equivalenze sono attive nell'implementazione SKOS, poiché non tutti gli strumenti collegati pubblicano dati in RDF. La BNCF scarica costantemente dati dagli strumenti collegati e disponibili nello SKOS da essi aggiornato; alcuni degli enti che li gestiscono, periodicamente, fanno l'inverso e ciò fa sì che si stabiliscano link reciproci, spesso esplicitati, come nel caso di LCSH in cui i link ai termini italiani del *Nuovo soggettario* sono segnalati anche con l'icona del nostro tricolore[18].
L'interoperablità con *Wikipedia* è evidentemente una delle più significative e si è sviluppata nell'ambito di un progetto di collaborazione con il capitolo italiano di

[16] Il riferimento è alla licenza Creative Commons: <http://thes.bncf.firenze.sbn.it/thes-dati.htm>.
[17] La sperimentazione è stata avviata (con la collaborazione di @CULT) tramite la creazione di un modello di apprendimento multidisciplinare a partire dalle tesi di dottorato di ricerca in versione elettronica depositate in BNCF. I risultati ottenuti sono tuttora in fase di analisi; margini di miglioramento potranno essere legati al raffinamento delle procedure di attribuzione di metadati sui documenti che compongono questo modello di apprendimento e dall'uso di algoritmi più sofisticati rispetto a quelli basati unicamente su leggi statistiche e frequenza delle parole.
[18] Sui collegamenti reciproci fra *Nuovo soggettario* e LCSH: Anna Lucarelli – Elisabetta Viti, *Florence–Washington Round Trip: Ways and Intersections between Semantic Indexing Tools in Different Languages*, «Cataloging & Classification Quarterly», 53 (2015), n. 3-4, p. 414-429 (pubblicato online il 19 Marzo 2015, <http://www.tandfonline.com/toc/wccq20/53/3-4>).

Wikimedia Foundation. Anche grazie all'impiego sperimentale di automatismi, i link con *Wikipedia* sono notevolmente aumentati; questa interoperabilità ha accresciuto le visite del Thesaurus e gli accessi all'Opac della BNCF. La biblioteca non ha imposto alcun obbligo di registrazione a chi scarica dati tramite il formato SKOS/RDF del Thesaurus, al fine di renderlo usabile il più facilmente possibile, una politica che si è confermata con la possibilità recentemente offerta di interrogare i dati tramite un endpoint SPARQL su Datahub che utilizza CKAN come piattaforma pubblica[19] (Fig. 3).

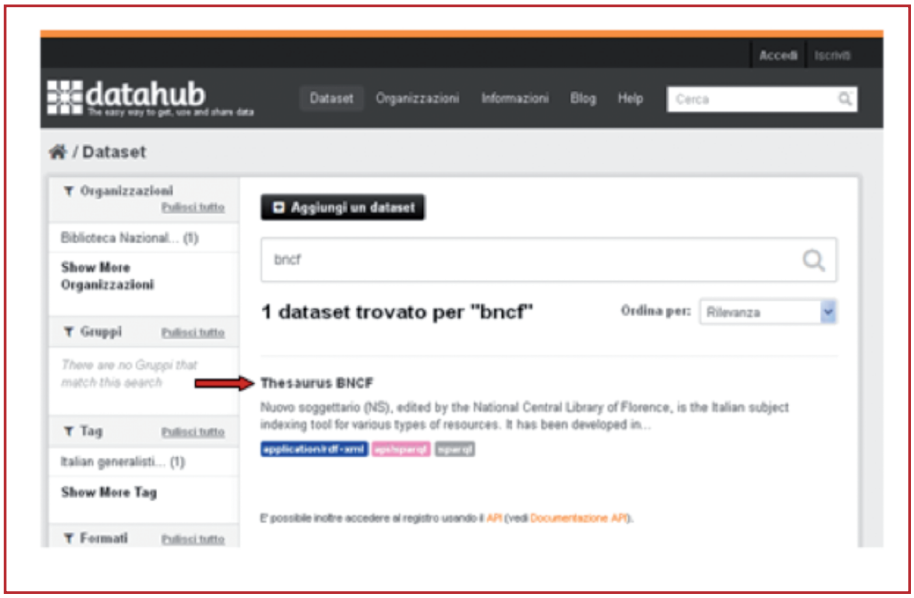


Figura 3: *Il data set del Thesaurus tramite la pagina di Datahub*

Pubblicazione dei dati e trasparenza informativa

I dati aperti del *Nuovo soggettario*, dunque, attivano interoperabilità e favoriscono forme varie di riuso. È interessante analizzare gli accessi al server, sapere quanti e quali utenti utilizzano i dati in formato SKOS, anche mese per mese, anno per anno. Si scopre, ad esempio, che viene scaricato da biblioteche di tutti tipi (dalla Nazionale Marciana, alla Library of Congress, dalla Biblioteca del Congresso del Cile al Sistema bibliotecario del Canton Ticino), da università italiane e straniere, da archivi di Stato (ad esempio, quello di Milano), da soprintendenze archivistiche, da musei (ad esempio, dal Museo Galileo di Firenze, oppure dai Musées royaux d'art et d'histoire di Bruxelles), da commissioni dell'Unione europea, da enti pubblici di vario genere (il Ministero dell'interno, la Regione Emilia Romagna, la

[19] Knowledge Foundation, è un sistema molto usato da amministrazioni pubbliche (sia straniere che italiane) per la pubblicazione dei propri open data; per il suo sito ufficiale:< http://ckan.org/>.

Provincia di Ravenna, il Comune di Firenze ecc.), da enti, fondazioni, aziende (come CNR, Fondazione Bruno Kessler, Cineca, CSI Piemonte, Cedoc, Data-Management[20]), da aziende sanitarie locali. Il riutilizzo dei dati è spesso testimoniato da esplicite clausole di attribuzione. Ma per la BNCF è interessante analizzare anche le applicazioni che ne vengono fatte, avere riscontri concreti e avviare canali informativi e comunicativi, poiché anche nell'epoca degli open data le biblioteche devono continuare a lavorare, oltre che sul piano tecnico, anche su quello delle politiche. Le esperienze realizzate dovrebbero comunque essere maggiormente pubblicizzate. È sempre interessante entrare in contatto con la riutilizzazione fuori contesto di dati aperti prodotti da biblioteche. Quando si scopre che possono trovare una loro utilità in altre amministrazioni pubbliche e in ambienti anche molto diversi da quelli di origine, se ne resta sorprendentemente colpiti, come quando qualche anno fa su «Che futuro!», periodico online dedicato all'innovazione, Giovanni Menduni, allora dirigente del Comune di Firenze e coordinatore dei progetti di open data di quell'amministrazione, riferendosi all'integrazione dei dati del *Nuovo soggettario* con altri *data set* del Comune, raccontava *Come abbiamo ordinato gli open data di Firenze con l'aiuto di una biblioteca* (Fig. 4)[21].

Figura 4: *Articolo sull'esperienza di riuso degli open data del Nuovo soggettario da parte del Comune di Firenze*

[20] L'uso dei dati del *Nuovo soggettario* da parte di Data Management, nell'ambito di Sebina, è stato dichiarato in modo formale e ha dato luogo ad uno scambio con la BNCF, utile per alcuni test e verifiche.

[21] http://www.chefuturo.it/2012/07/come-abbiamo-ordinato-gli-open-data-di-firenze-con-laiuto-di-una-biblioteca/.

National Libraries around the world are working to make their resources available on the Web, beyond the frontiers of their OPACs. The aim is to offer not only works and editions online but also structured data quality, assigned during the cataloging and digitalizing stages, so that they can be interlinked and made even more useful via the semantic Web. The Linked Open Data are a new and extraordinary opportunity, and give a new relevance to the social and public roles of the National Libraries. The National Central Library of Florence has just begun its Open Data experience with the Nuovo Soggettario Thesaurus, a tool developed towards multilingualism and interoperable with the data of other public administrations and memory institutions.

L'ultima consultazione dei siti Web è avvenuta nel mese di dicembre 2015.

SISTEMA MODUS

L'esperienza LOD (Linked Open Data) del Museo Galileo

Stefano Casati – Tiziana Possemato
Museo Galileo - @CULT

La Biblioteca digitale del Museo Galileo è stata istituita nel 2004 per offrire un servizio di consultazione di opere rare e importanti per la comunità degli storici della scienza. L'analisi delle esigenze specifiche del Museo Galileo ha successivamente mutato questo obiettivo, sottolineando l'esigenza di un progetto molto più ambizioso. Questa nuova fase è stata finalizzata alla creazione di un sistema informativo adeguato per la raccolta e il miglioramento delle diverse tipologie di risorse digitali - immagini, testi, video, suoni, animazioni, ecc. In questo contesto si inserisce il progetto MINERV@, che realizza un dataset secondo i principi dei Linked Open Data dal titolo "Banca dati Museo Galileo: strumenti, libri, fotografie, documenti". ALIADA è il framework open source di pre-lavorazione dei dati utilizzato nel progetto MINERV@, capace di automatizzare l'intero processo di creazione e pubblicazione di linked open data indipendentemente dal formato di origine dei dati. L'adozione dei LOD richiede alcune considerazioni e valutazioni di fattori quali: la qualità dei risultati ottenuti, le competenze necessarie per la realizzazione di un progetto, i costi di realizzazione. Il progetto MINERV@ dovrebbe rappresentare un punto di partenza di un più ricco percorso evolutivo: connettere il Museo Galileo a una rete di operatori culturali in grado di dare nuovo slancio all'economia digitale e offrire risorse all'industria culturale e creativa.

Nel dinamico e fluttuante universo linguistico delle discipline inerenti ai Beni culturali il termine GLAM è diventato consuetudinario, tanto che possiamo legittimamente considerare l'acronimo anglosassone per Galleries Libraries Archives Museums come espressione di un nuovo paradigma, un approccio innovativo per la tutela e la diffusione del patrimonio culturale. Negli ultimi dieci anni, specialmente in virtù del sempre più decisivo apporto delle tecnologie digitali, le istituzioni culturali hanno avvertito fortemente la necessità di estendere su vasta scala e con modalità diverse l'accesso al proprio patrimonio, estendendo l'orizzonte conoscitivo attraverso il confronto e la collaborazione con altre realtà per proporre un nuovo concetto di collezione. La crescita di ampie e condivisibili infrastrutture conoscitive, l'esigenza di affrancarsi dall'isolamento e dai limiti caratteristici di molte strutture tradizionali del sapere sembrano incarnarsi perfettamente nella filosofia GLAM di cui OpenGLAM, GLAM-Wiki e l'istituzione del MAB[1]

[1] Su OpenGLAM si veda: <http://openglam.org/ e su GLAM-Wiki <https://outreach.wikimedia.org/wiki/GLAM>. Per informazioni su Musei Archivi Biblioteche (MAB), istituito nel 2011 da Associazione Italiana Biblioteche (AIB), Associazione Nazionale Archivistica Italiana (ANAI) e International Council of Museum - Comitato Nazionale Italiano (ICOM Italia), si rimanda a <http://www.mab-italia.org/index.php/musei-archivi-biblioteche/mab-italia>.

SISTEMA MODUS

sono alcune delle più interessanti espressioni. Questi promettenti scenari per lo sviluppo e la trasmissione della conoscenza trovano condizioni ideali in ambienti Smart, luoghi permeati da creatività e innovazione tecnologica. L'Agenda Digitale Europea detta linee guida e indirizzi per rafforzare questa tendenza[2] ed è ormai diffusa l'idea che il bene culturale, tramite itinerari e mappe intelligenti tracciate dalle nuove tecnologie, debba oltrepassare i confini della sua fruizione fisica tradizionale assumendo una vasta dimensione sociale.

Il Museo Galileo racchiude in sé le condizioni essenziali per essere collocato in uno spazio *smart* in ottica GLAM. Negli ultimi quindici anni l'istituzione fiorentina ha consolidato questa tendenza avviando numerose iniziative, come il progetto Smartmuseum[3] realizzato per offrire interattività con i contenuti culturali in formato digitale in modo da soddisfare maggiormente le esigenze dei visitatori di musei, e soprattutto istituendo il nuovo settore della Biblioteca digitale[4]. Nata nel 2004, per fornire un servizio di consultazione online di opere destinate alla comunità degli storici della scienza e della tecnica, la Biblioteca digitale del Museo Galileo ha seguito inizialmente un'impostazione finalizzata alla realizzazione di collezioni digitali tematiche, ricorrendo al catalogo della biblioteca e alle bibliografie specializzate prodotte internamente, come la *Bibliografia italiana di storia della scienza* e la *Bibliografia galileiana*[5]. Un obiettivo che ha permesso la costituzione di una raccolta significativa, non tanto per l'aspetto quantitativo in sé - attualmente sono consultabili circa 30.000 volumi online - quanto per la rarità e la specializzazione dei testi. Nonostante questo aspetto positivo, il modello iniziale, fortemente orientato alla digitalizzazione di libri e manoscritti, risultava però inadeguato alla realtà culturale del Museo Galileo, un'istituzione dotata di un patrimonio documentario vario e eterogeneo. La dizione Museo Galileo non deve trarre in inganno; intorno al Museo, dove si esercitano attività di conservazione, di museografia, ecc., orbitano infatti altri settori: un Istituto che promuove attività di ricerca e di divulgazione in ambito specialistico, una Biblioteca di ricerca dotata di un ingente fondo librario e attiva nella produzione di bibliografie, un Archivio con importanti fondi (periodo compreso dal Settecento al Novecento), un Laboratorio fotografico

[2] Per informazioni si veda: <http://ec.europa.eu/digital-agenda/ e http://eur-lex.europa.eu/legal-content/IT/TXT/?uri=URISERV:si0016>.

[3] Per informazioni e approfondimenti sul progetto SMARTMUSEUM - Cultural Heritage Knowledge Exchange Platform: http://brunelleschi.imss.fi.it/smartmuseum/web/iintro.html.

[4] Sulla Biblioteca digitale del Museo Galileo consultare il sito Web: <http://www.museogalileo.it/esplora/biblioteche/bibliotecadigitale.html>, oltre a: Stefano Casati, *La Biblioteca digitale del Museo Galileo*, Biblioteche oggi (genn.-febbr. 2015), p. 45-51 e Stefano Casati, Fabrizio Butini, *Some remarks about Museo Galileo's digital collections*, in: *Collaborative research practices and shared infrastructures for humanities computing: 2nd AIUCD Annual Conference, AIUCD 2013, Padua, Italy, 11-12 December 2013*, proceedings of revised papers, ed. Maristella Agosti and Francesca Tomasi, Padova, CLEUP, 2014, p. 143-148.

[5] Per approfondimenti sulle attività bibliografiche della Biblioteca del Museo Galileo si rimanda al sito Web: <http://www.museogalileo.it/esplora/biblioteche/biblioteca.html>.

SISTEMA MODUS

per la produzione di documentazione iconografica, un Laboratorio multimediale e un'unità Web dove sono sviluppate e applicate le nuove tecnologie dell'informazione e della comunicazione del patrimonio culturale.
Per rendere fruibile una raccolta documentaria composta da materiale librario, strumenti scientifici, fotografie e risorse multimediali, la Biblioteca digitale doveva necessariamente adottare un modello diverso da quello iniziale e ideare un sistema informativo per raccogliere e integrare l'insieme delle collezioni digitali tematiche. Così, dopo alcuni tentativi, in stretta collaborazione con la Biblioteca dell'Istituto, è stato realizzato un sistema integrato di dati e risorse digitali a cui l'utente può accedere consultando l'OPAC della biblioteca.
Nell'archivio cumulativo della Biblioteca digitale confluiscono dati iconografici, bibliografici e archivistici, dati relativi agli strumenti scientifici e quelli inerenti alle applicazioni multimediali[6]. In virtù della piena integrazione dei dati è stato possibile intrecciare percorsi conoscitivi diversi e offrire nuove "letture" agli utenti remoti. Eseguendo, ad esempio, una ricerca semplice per il termine "Telescopio", si ottengono 116 risultati suddivisi in varie tipologie di documenti: 58 testi a stampa, 49 materiale grafico, 7 manoscritto, 1 materiale cartografico, 1 oggetto. Anche *Galileo//thek@*, progetto che mira a raccogliere e presentare tramite un'innovativa federazione integrata di risorse digitali tutti i documenti digitali galileiani, esemplifica questo nuovo modo di concepire la biblioteca digitale. Un approccio che fornisce all'utente potenti e raffinati strumenti di ricerca e che, per essere adeguatamente realizzato, necessita del coinvolgimento di numerose figure professionali con competenze informatiche, biblioteconomiche, scientifiche. Oltre a biblioteche digitali tematiche sono stati proposti anche percorsi di lettura di tipo "induttivo", focalizzando l'attenzione su testi particolarmente significativi per la storia della scienza come, ad esempio, i *Saggi di naturali esperienze* curati da Lorenzo Magalotti nel 1666.
Il percorso di sviluppo della Biblioteca digitale, contraddistinto da un costante e progressivo processo di integrazione e condivisione di dati e da un metodo di lavoro formatosi intorno ai concetti di cooperazione e interoperabilità, sembra naturalmente orientato verso l'universo in espansione dei LOD (Linked Open Data). Per rafforzare e completare questa propensione rimane indispensabile il passaggio ai metadati semantici e alla loro elaborazione in un formato adatto alla logica delle macchine. La Biblioteca digitale del Museo Galileo, benché abbia da tempo mostrato interesse per il Web semantico[7], segue ancora la logica del Web dei docu-

[6] Il sito Web del Museo Galileo presenta anche biblioteche digitali e mostre virtuali con livello di integrazione dati minimo ma con percorsi di navigazione strutturati. Per accedere a queste collezioni digitali: <http://www.museogalileo.it/esplora/biblioteche/bibliotecadigitale.html>.

[7] Il Museo Galileo ha prestato attenzione e interesse all'evoluzione del Web semantico, ha partecipato inoltre al Workgroup 2.2 *Semantic and multilingual Interoperability* e ha fornito i propri metadati a Europeana; per ulteriori informazioni si rimanda a: <http://www.museogalileo.it/esplora/progetti/progettieuropei.html>.

SISTEMA MODUS

menti e la dimensione di riferimento è rappresentata prevalentemente dal catalogo locale.
Le precondizioni di accesso al Web semantico della Biblioteca digitale del Museo Galileo sono sinteticamente individuabili nella propensione verso:

- l'integrazione di prodotto (archivio integrato);
- l'integrazione prodotto/servizio (biblioteca specializzata, archivio, laboratori multimediale, fotografico e di restauro);
- la naturalizzazione di ambienti tecnologici (biblioteca digitale, museo virtuale, Portale Galileo);
- innovative forme di offerta (con graduale passaggio da "didattica" ad "apprendimento");
- innovative forme di comunicazione e fruizione (videoguida, progetto Wiki).

A tutto questo si aggiunge il posizionamento storico del Museo Galileo all'interno di una rete di collegamenti istituzionali internazionali con altri soggetti culturali.
In questo contesto si inserisce il progetto MINERV@[8] che ha sancito la collaborazione fra la società @CULT di Roma e il Museo Galileo.
Con il progetto è stato realizzato un dataset secondo i principi dei Linked Open Data dal titolo "Banca dati Museo Galileo: strumenti, libri, fotografie, documenti".
I dati selezionati per la conversione in LOD (e dunque in RDF – *Resource Description Framework* - il modello dati selezionato dal W3C per la codifica, lo scambio e il riutilizzo di metadati strutturati) fanno parte della Banca Dati Cumulativa del Museo Galileo e riguardano:

- il catalogo delle collezioni di strumenti e macchine esposte nelle sale museali;
- il catalogo della biblioteca, dei materiali di archivio e delle collezioni fotografiche;
- *la Bibliografia italiana di storia della scienza;*
- *la Bibliografia galileiana*.

I dati selezionati per il progetto sono eterogenei (provenienti da differenti fonti) e molto strutturati (in Marc 21 come formato di origine); di grande interesse, dunque, per un progetto di condivisione delle informazioni attraverso la conversione e la pubblicazione in Linked Open Data.
Il paradigma dei Linked Open Data suggerisce un insieme di buone pratiche e regole per il trattamento e la pubblicazione dei dati. Le fasi in cui si articola il workflow di lavorazione possono essere concatenate in modalità differenti ma sono riconducibili alle seguenti attività:

1. Definizione delle specifiche:
   - identificazione e analisi della fonte dei dati;

[8] Nasce in risposta al bando FILAS: "Lazio Open Data. Trasparenza, informazione, coinvolgimento per i cittadini, le imprese e le istituzioni".

   - disegno dell'URI;
   - definizione delle licenze di distribuzione e uso.
2. Modellazione: la scelta o la creazione di ontologie e vocabolari controllati.
3. Generazione dei dati in formato RDF, tramite i linguaggi di mappatura disponibili:
   - conversione del dato;
   - bonifica di quanto prodotto;
   - creazione di collegamenti con altri dati.
4. Pubblicazione dei dati in RDF:
   - pubblicazione del dataset su uno o più portali o piattaforme (es: CKAN, Comprehensive Knowledge Archive Network, Socrata, …) e dei relativi metadati;
   - generazione di sistema per aumentare la ricercabilità del dato.
5. Valorizzazione dei dati tramite la costruzione di strumenti semplici ed efficaci di utilizzo del dataset; per esempio, tramite la costruzione di interfacce per la fruizione.

Nel caso specifico di MINERV@, data la natura dei dati originari e le finalità progettuali, le attività eseguite sono state:

1. *Individuazione e selezione dei dati*: la selezione è stata realizzata tra i circa 250.000 record bibliografici e di authority provenienti dal Museo Galileo, scegliendo quelli attinenti la bibliografia galileiana, prodotti in formato MARC21 grazie all'attività di catalogazione realizzata negli anni all'interno della Biblioteca del Museo. Una fase successiva ha riguardato la scelta dei singoli elementi descrittivi (tag Marc) inclusi nel record da utilizzare nella conversione in RDF; questa soluzione è stata guidata dal principio dell'interesse comune: tutte le informazioni utili all'utenza finale (ricercatori), come il titolo, l'autore, l'editore, l'anno di pubblicazione, le informazioni aggiuntive identificative delle risorse.
2. *Analisi e modellazione:* ciascun elemento è stato analizzato per individuare l'uso fattone all'interno del record e del catalogo (il Museo Galileo ha fatto un uso standard del formato MARC21, ma con delle eccezioni – estensioni di utilizzo e personalizzazioni – dovute alla necessità di gestire materiali di differenti nature e tipologie con un unico standard, nato soprattutto per lo scambio dei dati bibliografici).
   - Scelta delle ontologie: la selezione delle ontologie, utilizzate per la mappatura e conversione dei record in RDF, è stata guidata dallo studio di progetti nazionali ed internazionali affini a MINERV@, e di area tematica simile a quella del progetto.
     Tra le ontologie analizzate e poi utilizzate ci sono quelle di seguito elencate:
   - DCMI Metadata Terms – lo schema dei dati Dublin Core, considerato da

molti la madre di tutti i metadati perché è stato il primo schema di metadati bibliografici ispirato al mondo delle biblioteche e perché continua ad avere un ruolo chiave nella descrizione delle risorse sul Web.

- RDF Schema – sviluppato dall'RDF Core Working Group come parte del W3C Semantic Web Activity (Activity Statement, Group Charter), ed utilizzato da importanti progetti di linked data.
- BIBO (Bibliographic Ontology) – impiegato da importanti istituzioni per progetti di creazione di set di linked data (Library of Congress e British Library), utilizza a sua volta alcuni elementi presi da altri schemi di metadati, tra cui il già citato Dublin Core. Interessante per l'ambito biblioteconomico è anche la parte dedicata al trattamento degli autori, così come lo status dei documenti (peer-reviewed, accettato, rifiutato), e la possibilità di trattare ISBN, ISSN ed altri numeri di controllo.
- FRBR (FRBR Core, FRBR Extended, FRBRoo, FRBRer Model) – questo modello concettuale è stato tra i primi ad essere espresso in RDF, ed ha avuto differenti versioni, alcune ufficialmente riconosciute dall'IFLA (FRBRer) ed altre non ufficiali realizzate da comunità di sviluppatori e conoscitori del modello quando il fenomeno dei linked data è esploso. La versione analizzata è stata quella ufficialmente sviluppata e dunque riconosciuta dall'IFLA FRBR Study Group ed inserita nell'Open Metadata Registry. L'ontologia FRBRoo è interessante soprattutto per la descrizione di informazioni museali, e quindi per progetti trasversali di pubblicazione dei dati culturali nel Web.
- RDA Elements – prodotto dal Dublin Core Metadata Initiative (DCMI) e dallo Joint Steering Committee for Development of RDA (JSC).
- FOAF (Friend of a Friend) – questa ontologia, nata per codificare informazioni sulle persone in ambito dei social network, è diventata poi il primo strumento di creazione dei metadati relativi a persone nel Web semantico. Si è diffusa anche nel mondo delle biblioteche e degli archivi come schema di metadati utilizzato per la codifica di informazioni sulle persone, i luoghi ad esse associati, le relazioni con altre persone.

3. *Produzione dell'output e sua validazione*: la mappatura dei dati è stata utilizzata per l'implementazione delle regole di conversione, che sono confluite nel framework di conversione dei dati in RDF/XML. Il file RDF prodotto è stato poi analizzato da un gruppo di esperti di catalogazione per verificare la correttezza formale e sintattica dell'output, e la corretta implementazione delle regole di applicazione delle ontologie selezionate.
4. *Pubblicazione*: il dataset "Banca dati Museo Galileo: strumenti, libri, fotografie, documenti" è stato pubblicato in DataHub[9] (il portale della Open

[9] http://datahub.io/organization/museogalileo.

Knowledge Foundation, un network mondiale no-profit dedicato all'apertura delle informazioni) e nel Portale OpenData della Regione Toscana[10]. Con questo ultimo passaggio i dati sono diventati accessibili, aperti, riutilizzabili.

Il framework[11] open source di pre-lavorazione del dato utilizzato nel progetto MINERV@, capace di automatizzare l'intero processo di creazione e pubblicazione di linked open data indipendentemente dal formato di origine dei dati, era stato precedentemente sviluppato dalla società @CULT ed è poi confluito nell'ambito del progetto europeo ALIADA (Automatic publication under LInked dAta paradigm of library DAta)[12].
L'universo di riferimento del progetto è multidominio: da una parte, c'è la dimensione dello sviluppo tecnologico informatico/informativo che caratterizza gli elementi costitutivi del progetto MINERV@; dall'atra, c'è la dimensione della ridefinizione del ruolo di gallerie, biblioteche, archivi e musei che vengono a costituirsi sempre più come nodi di una rete strutturata, quindi più ampia, di fornitori e creatori di cultura. Il progetto MINERV@, avendo coniugato le due dimensioni, rappresenta la necessaria sperimentazione di innovazioni, sempre più urgente per la gestione e fruizione dei beni culturali.
La creazione di un patrimonio collettivo e condiviso, non attraverso una collezione fisica onnicomprensiva ma attraverso una rete di relazioni tra basi di conoscenza distinte ma collegate, è un volano di sviluppo formidabile verso nuove forme di collaborazione e sviluppo nel mondo GLAM, che facilita la creazione di:

- *innovativi servizi* basati sul principio del libero accesso;
- *nuovi prodotti* mediante il riutilizzo creativo dell'esistente;
- *evoluti processi* di cooperazione tra attori esterni, incoraggiati a sviluppare servizi basati sul patrimonio condiviso di conoscenza.

In ogni caso, l'adozione dei LOD - sia per le loro caratteristiche, sia per la natura specialistica di una realtà culturale come il Museo Galileo - richiede alcune considerazioni e valutazioni di fattori quali:

- la *qualità dei risultati* ottenuti: l'utenza di riferimento del Museo Galileo è la comunità di storici della scienza; l'ambiente della ricerca necessita, quindi, di informazioni di granularità minima e di dati controllati e di qualità. Questa esigenza potrebbe essere non allineata con ciò che i LOD determinano: informazioni atomiche, moltiplicate dal meccanismo dell'*interlinking* e collegate a fonti esterne che pur logicamente attinenti potrebbero creare una dinamica di *information overload*. Questo rischio diminuisce grazie ad una specificità dei LOD: i dati strutturati in RDF sono leggibili, comprensibili e riutilizzabili da

[10] http://dati.toscana.it/dataset/banca-dati-museo-galileo-strumenti-libri-fotografie-documenti.
[11] http://www.atcult.it/prodotti/aliada.php.
[12] http://www.aliada-project.eu/.

computer; in questo modo i processi inferenziali, ossia di deduzione e creazione di nuova conoscenza, possono esser compiuti da agenti automatici in grado di processare moli di dati in tempi che all'uomo non sarebbero possibili.

– Le *competenze necessarie* per la realizzazione di un progetto: la creazione e pubblicazione di dataset in linked open data, così come il riutilizzo del dataset prodotto, richiedono una competenza ancora non largamente diffusa e condivisa, soprattutto nell'aspetto di scelta ed utilizzo delle tecnologie e degli strumenti al momento disponibili[13].
– I *costi di realizzazione* di un progetto: strettamente connessa alla problematica precedente, è quella dei costi da sostenere. Anche questo aspetto può essere di natura transitoria; in ambito Web semantico molto si sta producendo e numerosi iniziano ad essere i progetti legati alla produzione, pubblicazione e diffusione di LOD, sia in ambito nazionale che internazionale. L'attenzione dedicata a tali tematiche, a livello mondiale, è testimoniata dall'impegno con cui si cercano di formulare linee guida, strumenti e metodologie che, una volta affermatisi come standard, contribuiranno inevitabilmente alla riduzione dei costi cui, oggi, bisogna far fronte.

Una riflessione a parte merita la problematica dell'apertura e libera condivisione dei dati, con riferimento alla tipologia di licenza da utilizzare. In linea del tutto definitoria, la licenza è un documento con cui sono illustrati diritti e doveri dell'utente. Nel caso dei Linked Open Data, esse devono risultare compatibili con la definizione di Open Data della Open Knowledge Foundation (OKF): «dati che possono essere liberamente utilizzati, riutilizzati e ridistribuiti, con la sola limitazione – al massimo – della richiesta di attribuzione dell'autore e della redistribuzione allo stesso modo». Le licenze per il *copyleft* (gioco di parole coniato sul termine copyright), salvo casi eccezionali, prevedono, inoltre, per l'utente la possibilità di utilizzo anche con finalità commerciali. In realtà molto è stato fatto e diffuse sono le licenze ormai a disposizione. Le più comuni sono le licenze CC (*Creative Commons*), affermatesi come standard *de facto* a livello internazionale; ne esistono di diversi tipi, ognuna basata su specifiche esigenze, ma tutte consentono e facilitano un ampio riuso dei dati grazie ad un buon equilibrio tra permessi e vincoli, in linea con la Open Definition. Per il contesto italiano sono state sviluppate le IODL (Italian Open Data License), da parte di Formez PA, al fine di dare a tutte le amministrazioni uno strumento chiaro e certificato in grado di facilitare la diffusione e il riutilizzo delle informazioni. Un ulteriore passo avanti è stato compiuto con l'affermarsi del principio "Open by default", espressione con cui le fonti governative italiane hanno introdotto una sostanziale innovazione con l'art. 9 del Decreto legge 18 ottobre 2012, n. 179, convertito con modificazioni dalla Legge 17 di-

[13] Come riferimento generale per le tecnologie utilizzate nella realizzazione del Web Semantic si veda: <https://en.wikipedia.org/wiki/Semantic_Web_Stack>.

SISTEMA MODUS

cembre 2012, n. 221: «*A partire dal 18 marzo 2013, scadenza dei novanta giorni previsti dalla Legge, dati e documenti pubblicati online dalle amministrazioni titolari – senza un'esplicita licenza d'uso che ne definisca le possibilità e i limiti di utilizzo – sono da intendersi come dati aperti*».

Il settore pubblico ha iniziato questo percorso virtuoso di apertura dei dati, all'interno dei piani di sviluppo comunitari[14]; qualche ritardo è invece da registrarsi per quanto riguarda il mondo GLAM e la possibilità di sfruttare appieno i vantaggi offerti: il valore del riuso delle informazioni culturali risiede principalmente nella possibilità – da parte di un ente o istituzione – di arricchire le proprie risorse informative, migliorare i servizi offerti ed ottimizzare la propria posizione sul mercato di riferimento.

Da quanto detto emerge con evidenza quanto sia strategico adottare le tecnologie semantiche e i Linked Open Data in ambito "Beni Culturali".

Per questo, il dataset realizzato con il progetto MINERV@ non dovrebbe rappresentare un punto di arrivo ma di partenza di un percorso evolutivo: connettere il Museo Galileo ad una rete di operatori culturali, secondo logiche di internazionalizzazione, in grado di dare nuovo slancio all'economia digitale e offrire risorse all'industria culturale e creativa. Quello che auspichiamo è la realizzazione di un sistema che permetta agli enti di condividere in modo proficuo e funzionale le reciproche conoscenze e competenze e di relazionarsi in modo costruttivo e cooperativo al fine di accrescere visibilità, capacità innovativa e competitività.

The Museo Galileo's Digital Library (MGDL) began in 2004, with the purpose of offering an online consultation service of rare and important historical scientific works to scholars, by turning specialized bibliographies into dedicated digital libraries. The analysis of specific needs of the Museo Galileo soon brought about a change of the MGDL's primary goal and highlighted the demand for a much more ambitious project. This new phase was aimed at creating an information system suitable for collecting and improving the different typologies of digital resources – images, texts, videos, sounds, animations, etc. In this context the project MINERV@ was launched, it produces a data set according to the principles of Linked Open Data, named "Banca dati Museo Galileo: strumenti, libri, fotografie, documenti". ALIADA is the open source framework for data processing, used in the project MINERV@; it's able to automate the whole process of creating and publishing linked open data regardless of the original data format. The adoption of the LOD requires considerations and evaluations of such factors as: the quality of the results obtained, the skills needed for the realization of a project, the costs of implementation. The project MINERV@ should be a starting point of a richer evolutionary path: to connect the Museo Galileo in a network of cultural operators able to give new impetus to digital economy and to provide resources to cultural and creative industry.

[14] Direttiva 2003/98/CE e sua modifica – 2013/37/UE.

L'ultima consultazione dei siti Web è avvenuta nel mese di dicembre 2015.

SISTEMA MODUS

Nuove prospettive per il Web archiving: gli standard ISO 28500 (formato WARC) e ISO/TR 14873 sulla qualità del Web archiving

Stefano Allegrezza
Università degli Studi di Udine

Il Web archiving è un argomento di forte attualità in quanto, come è noto, se non si individuano in breve tempo soluzioni efficaci e sostenibili nel lungo periodo, si rischia di perdere per sempre quello che si è prodotto e pubblicato sul Web negli ultimi venti-trenta anni, dal momento che tale materiale è caratterizzato da un'estrema mutevolezza e dinamicità e spesso interi siti Web cambiano o scompaiono nel giro di poco tempo. Le soluzioni che sono state proposte fino ad oggi sono parziali e non sempre hanno raggiunto l'obiettivo. Tuttavia, recentemente ci sono state due novità che sembrerebbero poter assicurare prospettive migliori: si tratta da una parte della proposta di un formato elettronico specificatamente pensato per l'archiviazione del Web (il formato WARC), dall'altra della pubblicazione di una specifica norma ISO dedicata alla qualità nella conservazione del Web (ISO/TR 14873:2013). La rilevanza dell'argomento per il settore dei beni culturali è tale che è opportuno fare un po' di chiarezza su queste tematiche analizzando sia lo stato dell'arte che le prospettive future.

Introduzione

Le questioni legate al *Web archiving* costituiscono un argomento di grande attualità dal momento che si rischia di perdere per sempre buona parte di quanto il Web ha prodotto negli ultimi decenni se non si individuano in breve tempo soluzioni efficaci e sostenibili nel lungo periodo. Infatti, il Web ha una natura effimera ed è caratterizzato da una grande mutevolezza e dinamicità; non è raro che interi siti cambino radicalmente o scompaiano nel giro di poco tempo[1], spesso senza lasciare alcuna traccia di sé. Alcuni contenuti, specialmente quelli presenti

[1] Diversi studi hanno esaminato la vita media delle pagine Web, con risultati che vanno da 44 giorni a 75 o 100 giorni. Si veda, tra gli altri: Brewster Kahle, *Preserving the Internet*, «Scientific American Special Report», 1998, <http://web.archive.org/web/19980627072808/http://www.sciam.com/0397issue/0397kahle.html>; Rick Weiss, *On the Web, Research Work Proves Ephemeral*, «Washington Post», 23 novembre 2003, <http://faculty.missouri.edu/~glaserr/205f03/Article_WebPub.html>.

SISTEMA MODUS

sui siti di news o sui social media, mutano con una rapidità tale da mandare in crisi qualsiasi strategia di archiviazione[2].

Coscienti di questa minaccia, le istituzioni – specialmente quelle che operano nel settore culturale – hanno effettuato investimenti notevoli per sviluppare gli strumenti e le tecniche necessarie per supportare soluzioni di *Web archiving* su larga scala, con l'obiettivo di trovare le modalità più adeguate per raccogliere e conservare il materiale pubblicato in maniera sempre più imponente sul World Wide Web. Le soluzioni che sono state proposte fino ad oggi sono parziali e non sempre hanno raggiunto l'obiettivo. Tuttavia, recentemente l'attenzione si è focalizzata su due proposte che sembrerebbero poter fornire nuove soluzioni: si tratta da una parte dell'individuazione di un formato specificatamente pensato per l'archiviazione del Web (il formato WARC), dall'altra della pubblicazione di una specifica norma ISO dedicata alla qualità nell'archiviazione del Web (ISO/TR 14873:2013). La rilevanza dell'argomento per il settore dei beni culturali è tale che è opportuno fare chiarezza su queste tematiche analizzando sia lo stato dell'arte che le prospettive future.

Lo standard ISO 28500:2009 (formato WARC)

Fino a qualche anno fa gli strumenti software utilizzati per effettuare "catture" del Web (i cosiddetti *Web crawler*, detti anche semplicemente *crawler*, *spider* o *robot*)[3] organizzavano le informazioni raccolte codificandole nel modo più disparato ed il più delle volte utilizzando formati proprietari. Per questo si è ben presto sentita l'esigenza di individuare un formato che costituisse un riferimento semplice e nello stesso tempo affidabile da utilizzare per la cattura dei contenuti Web, per la loro gestione e per l'interscambio di tali materiali tra organizzazioni diverse. Il formato WARC (Web ARChive), riconosciuto ufficialmente standard ISO il 15 maggio 2009 con la pubblicazione della norma ISO 28500:2009[4] "*Information and documentation — WARC file format*", risponde in pieno a queste esigenze.

Questo formato costituisce un'evoluzione del precedente formato ARC creato nel 1996 da Brewster Kahle[5] e Mike Burner ed utilizzato fin da allora da Internet

[2] In alcuni casi, i contenuti diventano inaccessibili quando un sito viene riprogettato; in altri casi, il contenuto semplicemente scompare dal Web quando viene sostituito con informazioni più aggiornate, con la conseguenza che l'utente si trova davanti a link non funzionanti o al "classico" errore di tipo "404 page not found".

[3] Il *crawler* è essenzialmente un *browser* con funzionalità minime e privo dell'interfaccia grafica, che ha il compito di "simulare" il comportamento di un utente che vuole visitare, ad esempio, una serie di siti Web. Esso prende in input una lista di URL (Uniform Resource Locator) di pagine Web e le visita una per una seguendone anche i link. Tra i *crawler* più conosciuti vanno citati Heritrix, HTTrack e Googlebot.

[4] L'ultima revisione dello standard risale al 2014.

[5] Brewster Kahle è il fondatore di Internet Archive, la biblioteca digitale *non profit* che ha lo scopo dichiarato di consentire un "accesso universale alla conoscenza". Tra le collezioni di materiali

SISTEMA MODUS

Archive e da numerose istituzioni culturali per archiviare il materiale – costituito dalle pagine Web e dai link di collegamento tra di esse – raccolto dai *crawler* durante le loro esplorazioni del Web (*Web crawls*). Alla base di questa estensione vi sono motivazioni riconducibili alle esperienze delle organizzazioni che facevano parte dell'International Internet Preservation Consortium (IIPC)[6]; esse stavano sperimentando difficoltà via via maggiori nelle operazioni di archiviazione e di gestione del crescente volume di informazioni prelevate da Internet; per questo il gruppo di lavoro sugli standard dell'IIPC sottomise al Working Group 12 del Sottocomitato 4 del Comitato Tecnico TC 46 dell'ISO (in breve: ISO TC46/SC4/WG12) un progetto denominato "WARC file format". Il progetto venne accolto dall'ISO come New Working Project (NWP) nel maggio 2005. Da allora tutte le fasi del processo di standardizzazione si sono rapidamente succedute fino a giungere, il 15 maggio 2009, alla pubblicazione delle specifiche del formato come standard ufficiale ISO[7].

Le specifiche del formato WARC stabiliscono le modalità attraverso le quali è possibile concatenare gli oggetti digitali che tipicamente costituiscono una pagina Web o un intero sito Web ed i relativi metadati in un unico file d'archivio, il file WARC appunto. In questo senso il formato può essere utilizzato per creare applicazioni per la "raccolta" (*harvesting*), la gestione, l'accesso e lo scambio dei contenuti Web tra le varie istituzioni archivistiche. Inoltre, i file WARC possono essere utilizzati per archiviare qualsiasi genere di contenuto digitale, recuperato sia attraverso il protocollo Hypertext Transfer Protocol (HTTP) che mediante altri protocolli, come il File Transfer Protocol (FTP).

Rispetto al precedente formato ARC, il formato WARC possiede specifiche definite in maniera più puntuale, è più flessibile ed offre nuove funzionalità, come la memorizzazione delle richieste HTTP, la possibilità di inserire metadati arbitrari, l'assegnazione di un identificativo per ogni file trovato, la gestione dei duplicati e delle migrazioni dei record e la segmentazione dei contenuti raccolti su più record.

digitali include anche i siti Web. L'interfaccia Web utilizzata da Internet Archive per la navigazione sui siti Web che sono stati archiviati nei vari istanti della loro vita (mediante altrettanti *snapshot*) dal software di acquisizione prende il nome di *Wayback Machine*.

6 L'International Internet Preservation Consortium è un consorzio internazionale il cui obiettivo principale è quello di acquisire, rendere disponibili e conservare per le generazioni future le informazioni raccolte da Internet. Costituitosi nel luglio 2004 per coordinare gli sforzi necessari per preservare i contenuti Internet per il futuro, conta oggi un numero elevato di istituzioni (biblioteche, archivi, musei e altre organizzazioni culturali) di vari paesi del mondo. Il sito Web del consorzio è raggiungibile all'indirizzo: <http://netpreserve.org>.

7 Il formato è attualmente mantenuto dal Sottocomitato 4 – Interoperabilità tecnica del Comitato tecnico ISO/TC 46 – Informazione e documentazione.

SISTEMA MODUS

La struttura di un file WARC

Il formato WARC appartiene alla categoria dei formati di tipo "contenitore" e la struttura di un file WARC è piuttosto elementare: è costituito da una semplice successione di record (Fig. 1) il primo dei quali di solito è una semplice descrizione dei record seguenti.

Ciascun record è composto da un *text header* seguito da un *content block*, che costituisce il contenuto vero e proprio. Il *text header* è costituito da una prima linea che dichiara la conformità a una specifica versione del formato (ad esempio, WARC/1.0) seguita da un certo numero di campi del tipo "nome:valore" che servono per fornire varie informazioni sul record (come l'URI del sito oggetto della cattura, la data di cattura ecc.); il tutto è concluso con una riga vuota che serve da separatore prima del blocco del contenuto. Come separatore tra un record e il successivo vengono utilizzate due linee vuote. Il *content block* è costituito dal risultato di un tentativo di recupero dal Web (pagine Web, immagini, informazioni di reindirizzamento degli URL, risultati della ricerca degli host DNS ecc.) o da altri dati (ad esempio, metadati o contenuti trasformati) che forniscono ulteriori informazioni sui contenuti archiviati.

A seconda del tipo di contenuto, i record possono appartenere ad una delle seguenti otto categorie: *warcinfo, response, resource, request, metadata, revisit, conversion e continuation*; il tipo di record viene specificato nel campo WARC-Type. Ad esempio, viene utilizzato un record di tipo *request* per memorizzare le informazioni inviate al server Web che ospita un determinato sito nella fase di richiesta di una determinata pagina Web; le risposte ottenute vengono memorizzate

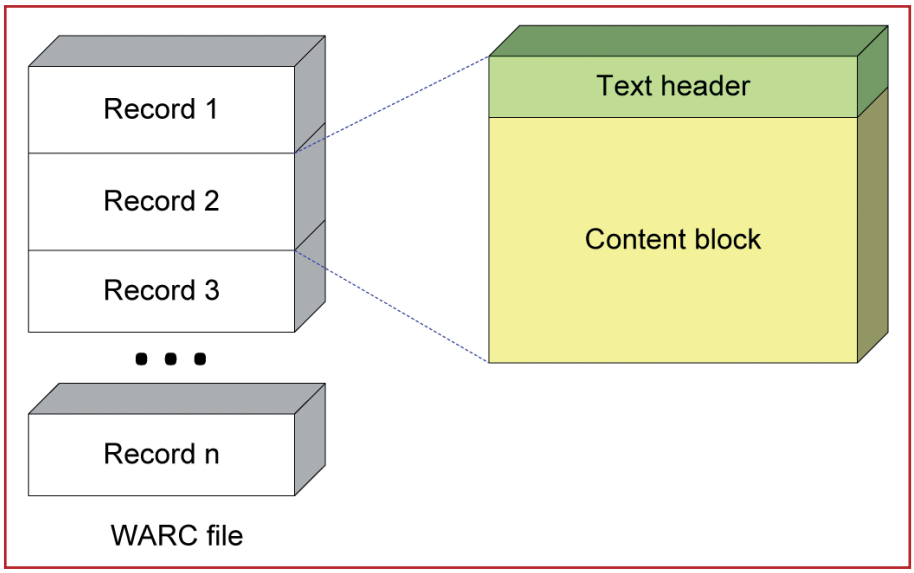


Figura 1: *La struttura di un file WARC*

in un successivo record di tipo *response*. Una immagine catturata dal Web costituisce il *content block* di un record di tipo *resource* (Fig. 2).

```
WARC/1.0
WARC-Type: response
WARC-Target-URI: http://www.digitalia.org/images/immagine.jpg
WARC-Warcinfo-ID: <urn:uuid:d7ae5c10-e6b3-4d27-967d-34780c58ba39>
WARC-Date: 2015-06-12T16:15:20Z
WARC-Block-Digest: sha1:UZY6ND6CCHXETFVJD2MSS7ZENMWF7KQ2
WARC-Payload-Digest: sha1:CCHXETFVJD2MUZY6ND6SS7ZENMWF7KQ2
WARC-IP-Address: 193.141.213.58
WARC-Record-ID: <urn:uuid:92283950-ef2f-4d72-b224-f54c6ec90bb0>
Content-Type: application/http;msgtype=response
WARC-Identified-Payload-Type: image/jpeg
Content-Length: 1902

HTTP/1.1 200 OK
Date: Tue, 19 Sep 2015 17:18:40 GMT
Server: Apache/2.2.31
Last-Modified: Mon, 16 Jun 2013 22:28:51 GMT
Accept-Ranges: bytes
Content-Length: 1662
Connection: close
Content-Type: image/jpeg

[image/jpeg – dati in binario]
```

Figura 2: *Esempio di record di tipo "response" (adattamento dallo standard ISO 28500:2009, Annex C "Examples of WARC records")*

La dimensione del file che si ottiene varia, evidentemente, a seconda dell'estensione del sito da catturare; la dimensione massima consigliata è intorno ad 1 GB, più che sufficiente per buona parte dei siti di piccole-medie dimensioni oggi esistenti. Nel caso in cui un sito Web abbia una dimensione maggiore è possibile suddividerlo in due o più file WARC.
Per il formato WARC è stato anche coniato un identificatore di tipo MIME; infatti, i file che si ottengono possiedono come MIME type la stringa *application/warc*.
Non rientra tra le finalità del presente articolo approfondire ulteriormente i dettagli tecnici del formato[8]; tuttavia, è importante metterne in evidenza la semplicità e la trasparenza[9], caratteristiche che costituiscono certamente un elemento a favore della possibilità di conservazione a lungo termine.

[8] Per un approfondimento sugli aspetti tecnici del formato si rimanda alla lettura della norma ISO 28500:2009.

[9] Per "trasparenza" di un formato elettronico si intende il grado di semplicità con cui è possibile fruire degli oggetti digitali codificati secondo quel formato, ad esempio utilizzando semplici strumenti di base.

SISTEMA MODUS

Analisi del formato WARC ai fini della conservazione a lungo termine

Come evidenziato da numerosi studi[10], un formato compatibile con un processo di conservazione digitale deve possedere innanzitutto le caratteristiche della non proprietà, dell'apertura, della standardizzazione e della trasparenza (Fig. 3).

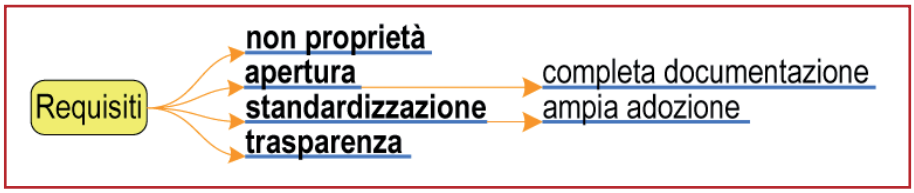


Figura 3: *I requisiti di base che un formato deve possedere per essere compatibile con un processo di conservazione a lungo termine*

Il formato WARC rispetta pienamente questi requisiti; infatti:

- è "non proprietario" in quanto è stato sviluppato sotto gli auspici dell'International Internet Preservation Consortium (IIPC); attualmente il gruppo di lavoro ISO responsabile del suo mantenimento è il TC46/SC4/WG12, che ha sede presso la Biblioteca nazionale francese[11];
- è "aperto" dal momento che le sue specifiche sono pubblicamente disponibili (si possono acquistare sul sito dell'International Standard Organization (ISO); le specifiche del formato nella versione *final draft* (ISO/FDIS 28500) che ha costituito la base per il processo di approvazione da parte dell'ISO sono, invece, liberamente disponibili[12]); inoltre, il formato è completamente documentato tramite le sue specifiche;
- è "standard" dal momento che è stato riconosciuto standard ISO 28500:2009; inoltre è ad oggi ampiamente adottato;
- è "trasparente", trattandosi di un formato che si limita a offrire un contenitore per gli oggetti digitali raccolti dal Web (che possono essere ovviamente nei formati più disparati).

Oltre a queste caratteristiche di base, il formato WARC gode di ulteriori interessanti proprietà[13]: non è sottoposto ad alcuna restrizione (in termini di licenze o brevetti); non vi sono meccanismi tecnici di protezione che potrebbero compro-

[10] Per un approfondimento si veda, tra gli altri, Stefano Allegrezza, *Requisiti e standard dei formati elettronici per la produzione di documenti informatici*, «Archivi & Computer», 19 (2009), n. 2-3, p. 42-82.

[11] Il sito Web di riferimento per le attività del gruppo è: <http://bibnum.bnf.fr/WARC/>.

[12] Il *final draft* è liberamente disponibile sul sito della Biblioteca nazionale francese all'indirizzo: <http://bibnum.bnf.fr/ WARC/WARC_ISO_28500_version1_latestdraft.pdf>.

[13] Si veda la sezione del sito della Library of Congress dedicata al formato WARC: <http://www.digitalpreservation.gov/ formats/fdd/fdd000236.shtml>.

metterne l'utilizzo in un prossimo futuro; infine, è auto-documentato (ciascuna risorsa – HTML, JPG, GIF ecc. – contenuta all'interno di un file WARC è corredata dalle informazioni su di essa).

Ormai tutti i più diffusi sistemi per il *Web archiving* hanno adottato il formato WARC per l'archiviazione dei materiali "catturati". Ad esempio, Heritrix[14], il *crawler* utilizzato da Internet Archive e da numerose istituzioni culturali (Fig. 4), adotta il WARC come formato predefinito per le sue "catture".

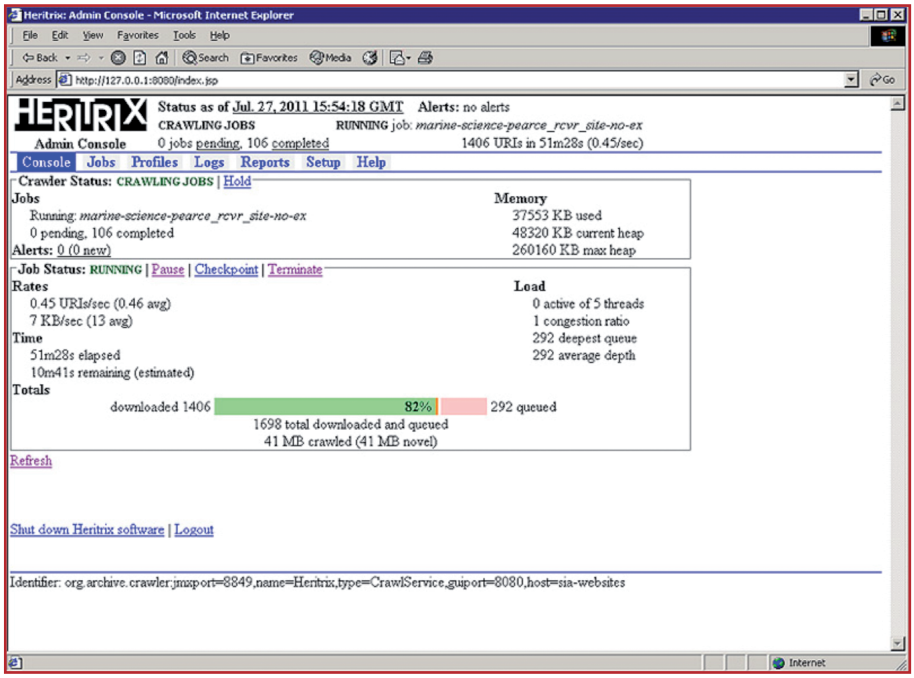


Figura 4: *La schermata di amministrazione di Heritrix*

Per la fruizione dei siti Web archiviati nel formato WARC sono disponibili estensioni per i più comuni browser e numerosi applicativi[15]. Tra questi merita sicuramente una segnalazione WERA (Web ARchive Access)[16], una soluzione il cui sviluppo è stato sponsorizzato dall'International Internet Preservation Consortium (IIPC) ed è

[14] Heritrix è il *crawler open source* sviluppato nell'ambito del progetto gestito da Internet Archive; è utilizzato per la raccolta dei siti Web da parte di numerose istituzioni, come Austrian National Library, Bibliothèque Nationale de France, British Library, Web Archiving Service della California Digital Library, Internet Memory Foundation, Library and Archives Canada, Library of Congress, National Library of New Zealand, Koninklijke Bibliotheek- Nationale Library van Nederland.

[15] Oltre all'applicativo WERA citato nel seguito, vanno segnalati i WARC tools per la gestione e lo scambio dei file WARC, la Wayback Machine di Internet Archive, NutchWAX e altri strumenti di ricerca che consentono l'accesso ai materiali raccolti.

[16] Il sito Web di riferimento è: <http://archive-access.sourceforge.net/projects/wera>.

SISTEMA MODUS

liberamente disponibile per chiunque voglia utilizzarla. Essa consente la navigazione e la ricerca nelle collezioni di siti Web archiviati e funziona in maniera analoga alla *Wayback Machine* dell'Internet Archive (ma a differenza di quest'ultima consente anche la ricerca *full-text* all'interno dell'archivio Web).

Lo standard ISO/TR 14873:2013

Per rispondere alla necessità di disporre di linee guida affidabili sulla gestione e valutazione delle attività di *Web archiving*, il 1 dicembre 2013 è stata pubblicata la norma ISO/TR 14873:2013 *Information and documentation — Statistics and quality issues for Web archiving*[17].

TECHNICAL REPORT

ISO/TR 14873

First edition 2013-12-01

Information and documentation — Statistics and quality issues for web archiving

Information et documentation — Statistiques et indicateurs de qualité pour l'archivage du web

ISO

Reference number ISO/TR 14873:2013(E)

© ISO 2013

Figura 5: *Il frontespizio della norma ISO/TR14873:2013*

Si tratta di un notevole passo avanti nel tentativo di standardizzare le metodiche e le tecniche che sono alla base del *Web archiving*. La norma si concentra essenzialmente sui principi e metodi del *Web archiving* e ne stabilisce la terminologia, gli indicatori statistici da rilevare e i criteri di qualità, tenendo in debita considerazione le necessità di una vasta gamma di utenti (biblioteche, archivi, musei, centri di ricerca ed istituzioni culturali). Ad esempio, la terminologia utilizzata tenta di riflettere la vasta gamma di interessi e competenze, cercando un equilibrio tra archivistica, informatica, management e biblioteconomia. È destinata in primo luogo ai professionisti direttamente coinvolti nell'archiviazione del Web ma è utile anche agli organismi di finanziamento delle istituzioni culturali e altri *stakeholders*.

La norma ISO/TR 14873:2013, composta da 54 pagine, è suddivisa in sei sezioni. Dopo le consuete sezioni iniziali costituite dal campo di applicazione (*Scope*) e dalle definizioni (*Terms and definitions*), la terza sezione (*Methods and purposes of Web archiving*) prende in esame i metodi e gli obiettivi del *Web archiving*, analizzandone i meccanismi di raccolta, le modalità di accesso e di descrizione, le stra-

[17] Il sottocomitato tecnico dell'ISO che ha curato e mantiene lo standard è l'ISO/TC 46/SC 8 *Quality - Statistics and performance evaluation*. La norma andrebbe letta in congiunzione con le altre due norme ISO 2789:2013 *Information and documentation – International library statistics* e ISO 11620:2014 *Information and documentation – Library performance indicators*.

SISTEMA MODUS

tegie di conservazione, gli aspetti legali e le motivazioni.
La quarta sezione (*Statistics*) prende in esame tutta una serie di indicatori statistici che possono essere utilizzati per effettuare misurazioni quantitative e qualitative. Ad esempio, per misurare la dimensione di un archivio Web la norma suggerisce tre possibili indicatori: il conteggio degli URL; il conteggio dei domini o degli host; il conteggio dei byte.
Vi sono poi indicatori di tipo statistico che affrontano altri aspetti del problema; di particolare interesse sono quelli relativi alla *digital preservation*, che prendono in esame questioni come la percentuale di risorse che sono state duplicate, la percentuale di risorse perse o deteriorate, la percentuale di risorse con un formato di file ben identificato, la percentuale di formati per i quali è stata individuata una ben precisa strategia di conservazione ed, infine, la percentuale di risorse che sono state sottoposte ad un controllo antivirus.
Infine, l'ultima sezione (*Usage and benefits*) mette in evidenza i benefici che possono derivare dall'applicazione della norma a tutte le possibili figure professionali coinvolte a vario titolo nella attività connesse con il *Web archiving* (archivisti, *digital curators*, informatici, manager, etc.).
Il vantaggio innegabile che si deve riconoscere a questo standard è quello di essere riuscito a conferire sistematicità a tutto un insieme di informazioni spesso eterogenee e, fino ad ora, male organizzate.

Prospettive future

Sebbene la fase di archiviazione del Web sia di importanza fondamentale in una strategia di conservazione a lungo termine, è evidente che da sola non esaurisce il problema della conservazione, dal momento che occorre anche mantenere nel tempo l'accessibilità alle risorse raccolte e quindi affrontare tutta una serie di problemi che riguardano l'obsolescenza dei formati, quella dei supporti dove le informazioni raccolte vengono archiviate, il mantenimento delle relazioni tra i vari oggetti digitali ecc.
Oltre a ciò occorre tenere presente alcune questioni che rimangono tutt'ora irrisolte:
– innanzitutto, la sterminata estensione del Web rende di fatto impossibile effettuare una archiviazione esaustiva e, tantomeno, conservarlo nella sua interezza (anche facendo ricorso a sistemi informatici dotati di capacità di memorizzazione straordinarie);
– in secondo luogo, il Web è per sua stessa natura estremamente dinamico (i contenuti delle pagine Web dei siti cambiano continuamente e spesso vengono generati *on the fly* a seconda delle richieste che giungono da parte degli utenti) ed è impensabile ricorrere a una soluzione che riesca a "scattare" – conservandone copia – continui *snapshot* di ciascun sito Web che si intende conservare e che muta istante dopo istante[18]; per questo è evidente che occorre necessariamente proce-

[18] Anche *Internet Archive* è in grado, con i suoi *crawler*, di raccogliere delle "istantanee" dei siti Web sulla base di un programma temporale che tiene conto di diversi fattori, ma certamente non di

dere in modo selettivo, con criteri sia di tipo tematico che legati all'importanza che determinati siti possono avere per le future generazioni a seconda delle varie chiavi di lettura (storica, sociologica, politica, economica ecc.);
– infine, vi è l'oggettiva difficoltà – se non addirittura l'impossibilità – di accedere al cosiddetto *deep Web*[19], cioè a quella porzione del Web comunemente non indicizzata dai motori di ricerca e che, secondo alcune stime è 4.000-5.000 volte più grande del cosiddetto *surface Web*, cioè della parte accessibile, per un volume totale di oltre 10 Petabyte.

Nonostante tutte queste difficoltà, si comprende quanto la conservazione di queste informazioni sia di importanza fondamentale per la futura ricerca in ogni campo delle scienze e numerose organizzazioni a livello internazionale hanno cominciato ad occuparsi di *Web archiving*[20], ottenendo risultati incoraggianti anche grazie all'impiego del formato WARC.

Conclusioni

Per molto tempo il *Web archiving* ha costituito una sfida impari per gli archivisti e i *digital curators* e ciò ha portato alla perdita di innumerevoli contenuti Web, oggi non più accessibili né conservati da alcuna istituzione.
Il formato WARC, che è aperto, non proprietario, standard, ben documentato ed ampiamente adottato rappresenta indubbiamente un notevole passo avanti verso la soluzione del problema dell'archiviazione, consentendo la raccolta, la gestione e l'interscambio delle risorse "catturate" nel Web. Pur non potendo essere additato come la soluzione definitiva a tutte le criticità, costituisce comunque una eccellente risposta al problema dell'obsolescenza dei formati elettronici dal momento che consente di ricondurlo a quello di un unico formato. Inoltre, la sua standardizzazione offre una garanzia di durata nel tempo, assicura l'interoperabilità tra il patrimonio raccolto da istituzioni archivistiche diverse e contribuisce a far sì che il *Web archiving* diventi una delle attività che possono essere svolte con efficacia dalle organizzazioni che si occupano di conservazione del patrimonio culturale.

conservare una copia di ciascun sito Web istante per istante.

[19] Il *deep Web* è costituito da vari tipi di pagine, tra cui in particolare: *pagine dinamiche*, cioè costruite al momento a seguito di valori forniti dall'utente (ad esempio, il prezzo di un biglietto del treno determinato in un certo istante); *pagine scollegate*, cioè non accessibili tramite un percorso che parta dalla home del sito, ma solo a chi ne conosca l'indirizzo; *pagine private*, accessibili solo attraverso l'inserimento di credenziali di autenticazione (*username* e *password*); *pagine ad accesso limitato*, per esempio quelle il cui accesso è protetto dai CAPTCHA. Cfr. la sezione dedicata al Web archiving sul sito Conservazione Digitale del Centro di eccellenza italiano sulla conservazione digitale <http://www.conservazionedigitale.org/wp/approfondimenti/ web-archiving-2/>.

[20] A questo proposito occorre riconoscere che la situazione delle istituzioni culturali in Italia è in netto ritardo rispetto a quanto avviene negli altri Paesi, in special modo rispetto a quelli del mondo anglosassone.

Infine, la definizione della norma ISOTR 14873:2013 costituisce una risposta alla mancanza di linee guida affidabili sulle attività di gestione e valutazione del *Web archiving* e rappresenta un notevole passo avanti nel tentativo di standardizzare le metodiche e le tecniche che ne sono alla base.
Nel complesso si tratta di due novità che contribuiscono ad alimentare la speranza di riuscire a conservare per le generazioni future l'imponente mole di informazioni presenti sul Web che altrimenti rischierebbe di andare irrimediabilmente perduta.

Web archiving is a topic of great relevance because, if we do not identify quickly effective and sustainable solutions, we may lose what has been produced and published on the Web in the last twenty/thirty years. In fact, this material is characterized by extreme changeability and dynamism and often entire Websites change or disappear in a short time. The solutions that have been proposed up to now are partial and have not always achieved the goal. However, recently there are two developments that seem to provide new tools: the proposal of a file format specifically designed for Web storage (the WARC format) and the publication of a specific ISO standard devoted to quality in the preservation of the Web (ISO/TR 14873:2013). The relevance for the cultural heritage sector is such that it is important to go into depth on these news.

L'ultima consultazione dei siti Web è avvenuta nel mese di dicembre 2015.

SISTEMA MODUS

Riferimenti bibliografici

Burner – Kahle 1996 Mike Burner - Brewster Kahle. *ARC File Format*, 15 September 1996, <http://www.archive.org/Web/ researcher/ArcFileFormat.php>.

Hanzo Web Archiving 2014 Hanzo Web Archiving. ISO *28500 WARC The open standard for Web content preservation for eDiscovery*, disponibile previa registrazione sul sito: <http://www.hanzoarchives.com/learn/whitepapers/Web-archiving-iso-28500-warc-white-paper>.

ISO 28500: 2009 2013 *ISO 28500:2009. A new standard for the WARC file format*, «Engineer Live Magazine», 21st February 2013 <http://www.engineerlive.com/content/22130>.

Kahle 2008 Brewster Kahle. *Preserving the Internet*, «Scientific American Special Report», 1998, <http://Web.archive.org/Web/19980627072808/ http://www.sciam.com/0397issue/0397kahle.html>.

Kim – Ross 2012 Yunhyong Kim – Seamus Ross. *Digital Forensics Formats: Seeking a Digital Preservation Storage Container Format for Web Archiving, «International Journal of Digital Curation», 7(2012), n.2 <http://www.ijdc.net/index.php/ijdc/article/view/217/286>.*

Kunze 2005 John Kunze. *WARC: an archiving format for the Web*, 5th International Web Archiving Workshop (IWAW05), <http://www.iwaw.net/05/kunze.pdf>.

Niu 2012 Jinfang Niu. *Functionalities of Web Archives*, «D-Lib Magazine», 18 (2012), n. 3-4 <http://dlib.org/dlib/march12/niu/03niu2.html>.

Niu 2012 Jinfang Niu. *An Overview of Web Archiving*, «D-Lib Magazine», 18 (2012), n. 3-4 <http://dlib.org/dlib/march12/niu/03niu1.html>.

Pennock 2013 Maureen Pennock. *Web-Archiving: DPC Technology Watch Report 13-01*, Digital Preservation Coalition. 2013 <http://www.dpconline.org/component/docman/doc_download/865-dpctw13-01pdf-dpctw13-01pdf>.

Sinibaldi-Buongiorno 2012 Alessandro Sinibaldi – Paolo Bartolomeo Buongiorno. *Manuale di conservazione digitale*, Milano, Franco Angeli, 2012.

Thompson 2008 *Dave Thompson. Archiving Web Resources*, DCC Digital Curation Manual, [a cura di] Seamus Ross, Michael Day, (December 2008), <http://www.dcc.ac.uk/resource/curation-manual/chapters/Web-archiving>.

Uk National Archives 2011 Uk National Archives. *Web Archiving Guidance*, 2011, <http://nationalarchives.gov.uk/documents/ information-management/Web-archiving-guidance.pdf>.

SISTEMA MODUS

Web archiving 2006 *Web archiving, [a cura di] Julien Masanès. Berlin:* Springer, 2006.

Weiss 2003 *Rick Weiss. On the Web, Research Work Proves Ephemeral*, Washington Post, 23 novembre 2003, <http://faculty.missouri.edu/~glaserr/205f03/Article_WebPub.html>.

Sitografia

The WARC File Format (ISO 28500) - Information, Maintenance, Drafts;sezione del sito della Bibliothèque Nationale de France dedicata la formato WARC, <http://bibnum.bnf.fr/WARC/>.

WARC File Format specifications; pagina dedicata agli Internet Archive ARC access tools su Sourceforge <http://archive-access.sourceforge.net/warc>.

Heritrix 3.0 and 3.1 User Guide, pagina dedicata all'utilizzo del formato WARC da parte del *crawler* Heritrix sul sito del progetto *open-source* Heritrix; si veda, in particolare, la sezione *: Output/WARC files* <https://Webarchive.jira.com/wiki/display/Heritrix/Heritrix +3.0+and+3.1+User+Guide>.

WARC (Web ARChive), pagina dedicata al formato WARC sul sito del progetto *open-source* Heritrix; si vedano in particolare gli esempi di file WARC <https://Webarchive.jira.com/wiki/pages/viewpage.action?pageId=4817>.

OpenWayback: General Overview (OpenWayback è un'applicazione open source Java finalizzata alla ricerca e all'accesso al material Web archiviato) <https://github.com/iipc/openwayback/wiki/General-overview>.

Sezione della Library of Congress Sustainability of Digital Formats Planning for Library of Congress Collections dedicata al formato WARC, <http://www.digitalpreservation.gov/formats/fdd/fdd000236.shtml>.

Sezione dedicata al Web archiving" sul sito Conservazione digitale del Centro di eccellenza italiano sulla conservazione digitale <http://www.conservazionedigitale.org/wp/approfondimenti/Web-archiving-2/>.

SISTEMA MODUS

Progetti

Linked data: il mondo di internet e il ruolo delle biblioteche, degli archivi e dei musei

Margherita Aste - Maria Cristina Mataloni - Luca Martinelli
ICCU

Introduzione

Il mondo dell'informazione in rete è sempre stato fortemente interessato al patrimonio di dati provenienti da biblioteche archivi e musei, una grande quantità di dati di qualità, fortemente strutturati e rispondenti a regole condivise. D'altro canto anche biblioteche, musei e archivi sono estremamente interessati all'integrazione dei propri dati in rete al fine di garantirne maggiore visibilità e riusabilità.
Numerose sono le esperienze di collaborazione tra biblioteche, musei e archivi a livello nazionale[1] e internazionale nell'ambito dell'International Federation of Library Associations (IFLA)[2] per il trattamento comune di patrimoni culturali diversi.
Le attività sono focalizzate alla definizione di standard e strumenti concettuali per la descrizione delle proprie risorse e a garantire l'interoperabilità e la navigazione tra le informazioni.
I modelli dei dati per la descrizione dei diversi tipi di materiali sono molto diversi tra loro.
In ambito bibliotecario è stato sviluppato il nuovo standard di descrizione delle risorse RDA[3] progettato per il mondo digitale e indirizzato non solo alle biblioteche ma anche ad altre istituzioni culturali, come musei e archivi.
Nel mondo archivistico la standardizzazione e l'uniformità del linguaggio sono caratteristiche piuttosto recenti.
Nell'ambito del settore museale è stato elaborato il modello concettuale per la descrizione degli "oggetti culturali" CIDOC-CRM[4].
Nel suo realizzarsi, il progetto che prevede la pubblicazione in Linked Open Data (LOD) degli archivi del Servizio Bibliotecario Nazionale (SBN) ha messo in luce le grandi potenzialità che tale formato può avere per la stessa base dati di SBN sia a

[1] MAB, acronimo per Musei Archivi e Biblioteche, coordinamento nazionale promosso nel 2011 da AIB, ANAI e ICOM Italia.
[2] http://www.ifla.org/node/8664.
[3] Resource Description & Access vedi http://www.rdatoolkit.org/.
[4] http://www.cidoc-crm.org.

livello di visibilità nei confronti degli utenti dei tradizionali motori di ricerca, ma anche per gli stessi bibliotecari, che avrebbero delle fonti autorevoli da utilizzare come utile ausilio alla catalogazione.
Il progetto ha altresì evidenziato alcuni aspetti problematici divenendo quindi spunto di ulteriore riflessione nel merito.
La pubblicazione in LOD dei dati SBN permetterebbe anche ad altre basi dati di potersi avvalere dell'autorevole lavoro svolto dai bibliotecari.
In questo senso, l'accordo sottoscritto nel marzo scorso con l'Associazione Wikimedia Italia e la presenza di un "wikipediano in residenza" presso l'ICCU stanno già delineando una collaborazione proficua fra l'Istituto, l'enciclopedia libera Wikipedia e gli altri progetti correlati, come il database di dati strutturati libero Wikidata.

1. La base dati SBN in Linked Open Data

Il catalogo SBN può prestarsi in maniera particolarmente efficace alla fruizione in modalità Linked Open Data. È per tale motivo che il Comitato Tecnico Scientifico per SBN, nel 2014, ha dato mandato all'ICCU di costituire un gruppo di lavoro per la definizione di un prototipo di sperimentazione dei dati del patrimonio informativo delle biblioteche della rete SBN in LOD, secondo le modalità del Web semantico, così come già sperimentato da analoghe istituzioni straniere come la BNF e il RISM.
Il Gruppo di lavoro, costituito da personale dell'ICCU e del VAST LAB PIN dell'Università di Prato, ha analizzato in prima battuta il set di dati SBN che maggiormente si prestava alla pubblicazione in LOD, decidendo di utilizzare i dati che non presentavano specificità inerenti la tipologia di materiale e rinviando la mappatura degli altri documenti (musicali, grafici, cartografici e audiovisivi).
L'architettura della base dati SBN è di tipo relazionale, ma l'archivio derivato che è visibile all'utenza dell'OPAC SBN perde la caratteristica relazionale, essendo esportato in formato UNIMARC.
Questa peculiarità, che potrebbe sembrare riduttiva, può rivelarsi invece utile all'approccio Web dal momento che il formato UNIMARC risponde ad uno standard internazionale (ISO2709) e viene utilizzato da diversi importanti Paesi (ad esempio la Francia).
La prima fase del progetto ha riguardato la mappatura dei campi UNIMARC utilizzati in SBN e, quindi, la scelta delle ontologie e del modello concettuale di riferimento.
Tale scelta è stata essenzialmente guidata dalla volontà di uniformarsi a modelli e standard internazionali che consentissero l'integrazione e lo scambio di informazioni bibliografiche tra i differenti settori della cultura e ne consentissero la strutturazione in RDF (Resource Description Framework)[5].

[5] Il Resource Description Framework (RDF) è uno standard flessibile proposto dal consorzio W3C per la codifica, lo scambio e il riutilizzo di metadati. Il data model è formato da risorse, presenti nel Web con un URI, proprietà ovvero relazioni e valori, anch'essi identificati da URI.

2. Analisi

Come accennato, l'analisi iniziale si è concentrata principalmente sulla scelta del modello concettuale e del tipo di ontologia da utilizzare, tenendo conto anche del fatto che una mappatura dei tag UNIMARC in RDF non è stata ancora mai realizzata. La scelta è caduta sui seguenti componenti:

- il CIDOC Conceptual Reference Model (CRM)[6], modello concettuale di riferimento, che costituisce la struttura formale per descrivere i concetti impliciti ed espliciti e le relazioni nella documentazione del patrimonio culturale;
- il Functional Requirements for Bibliographic Records object oriented (FRBRoo)[7], ontologia di riferimento che fornisce un sistema concettuale dotato di una complessa rete di classi e proprietà in grado di descrivere le principali entità e relazioni dell'ambito bibliografico, permettendo l'integrazione e lo scambio di dati di diversa tipologia e formato nell'ambito dei beni culturali.

Attraverso le classi e le relazioni di FRBRoo si arriva alla costruzione di *path* o "catene di relazioni" per mezzo delle quali è possibile esplicitare concetti anche complessi. Le operazioni di mappatura consistono essenzialmente nel creare di volta in volta corrispondenze fra ognuno dei campi UNIMARC (solitamente sintetici e semanticamente difficili da esportare al di fuori dell'ambito biblioteconomico) e uno o più *path* FRBRoo equivalenti. In questo modo si rende possibile la lettura semantica dell'oggetto bibliografico, che viene arricchito ulteriormente di relazioni a volte ricavate da altri dati standardizzati già esistenti sul Web (ad esempio Geonames, *authority file* online quali il VIAF, etc.).

La seconda fase di analisi ha riguardato la scelta della tipologia dei dati da trattare. Inizialmente si è posta attenzione alle notizie catalogate con un set minimo di dati comuni[8]; in una fase successiva sono stati mappati anche i dati specifici. Sulla base dei criteri citati sono stati selezionati 300 record relativi a documenti che descrivono monografie moderne, monografie antiche e periodici.

In relazione allo schema concettuale FRBR, va precisato che le informazioni presenti nello scarico UNIMARC SBN si riferiscono essenzialmente alla Manifestazione

[6] Il CIDOC CRM è il risultato dell'attività svolta per oltre 10 anni da parte dei gruppi di lavoro costituiti nel CIDOC in seno al CIDOC ICOM's International Committee for Documentation. Dal 12 settembre 2006, il CIDOC CRM è riconosciuto come standard ISO 21127:2006 e nel dicembre 2014 è stato aggiornato nella nuova versione ISO 21127:2014.

[7] L'ontologia FRBRoo è nata nel 2003 dal gruppo di lavoro FRBR/CIDOC CRM, che ha avuto lo scopo di allineare i due modelli di dati utilizzati in ambito bibliotecario (FRBR) e museale (CIDOC CRM) al fine di garantirne l'interoperabilità semantica.

[8] Per set minimo di dati comuni in SBN si intende la descrizione di una notizia bibliografica con i soli dati essenziali per l'identificazione del documento senza entrare nel merito delle caratteristiche catalografiche più specifiche di ciascuna tipologia di materiale. Ad esempio, si può descrivere una diapositiva riportando il titolo, l'autore, i dati di pubblicazione senza dare indicazioni sul colore, dimensione, ecc.

del documento, essendo praticamente prive di riferimento all'Opera, all'Espressione e ai dati gestionali, tipici questi ultimi del livello Item e presenti nelle basi dati locali.
Per quanto riguarda gli archivi di autorità, al momento viene esportato in UNIMARC solo quello relativo agli autori, l'unico che è stato quindi mappato.

3. Mappatura

Il lavoro di mappatura è stato realizzato a partire dal tracciato del record UNIMARC che si presenta strutturato in campi, sottocampi e indicatori, con l'esclusione, in un primo momento, dei tag relativi alle specificità di Grafica, Cartografia, Musica e Audiovisivi e di quelli non utilizzati in SBN.
Nelle operazioni di mappatura, effettuate utilizzando il documento "FRBR object-oriented definition and mapping to FRBR-ER" (version 0.9 draft)[9], sono state create le corrispondenze fra campi, sottocampi e indicatori UNIMARC e uno o più *path* FRBRoo.
Ciò ha permesso di sviluppare le procedure di codifica in linguaggio RDF delle informazioni presenti nel catalogo SBN, rendendo i dati in formato *machine readable* e subito condivisibili e consultabili sul Web.
Tra l'altro, diversi campi mappati hanno il loro corrispondente valore codificato in tabelle, alcune ad uso interno, altre compatibili con gli standard UNIMARC.

4. Normalizzazione, arricchimento e conversione dei dati

4.1. Normalizzazione dei dati

La conversione dei dati UNIMARC in triple RDF ha richiesto una fase di normalizzazione mirata all'analisi della coerenza e correttezza delle informazioni.
La qualità del dato in entrata rappresenta un aspetto fondamentale durante la fase di conversione, perché un dataset con valori non corretti può rendere inefficienti o addirittura impossibili alcune operazioni di confronto, di similitudine e di allineamento dei dati, anche al fine di possibili link con database esterni.
La base dati SBN presenta un numero elevato di record in cui molti campi sono inseriti in modalità testo libero che mal si prestano ad una normalizzazione automatizzata; va considerato, inoltre, che le numerose importazioni di dati avvenute in maniera automatica provenienti da archivi non SBN hanno a volte compromesso l'uniformità dei dati. Questi problemi andrebbero analizzati e approfonditi ulteriormente, sia ai fini dei collegamenti con risorse esterne, sia per consentire un miglior risultato nelle ricerche mirate dell'utente finale. In particolare, va sottolineato che la produzione di LOD di qualità dipende da un buon livello di pulizia dei dati del catalogo SBN.

[9] Il documento è disponibile al link http://www.ifla.org/files/assets/cataloguing/frbrrg/frbr-oo-v9.1_pr.pdf.

4.2. Arricchimento dei dati

Una delle maggiori potenzialità offerte dai Linked Open Data è la possibilità di ampliare i propri dati in ricchezza di contenuto, quantità e qualità grazie alle informazioni che possono essere associate, attraverso un collegamento stabile, a fonti esterne autorevoli che possiedono già una descrizione accurata di numerose entità. Per tale motivo è indispensabile scegliere con particolare attenzione un *Uniform Resource Identifier* (URI), ossia una stringa alfanumerica *stabile* che identifichi *univocamente* un determinato oggetto della base dati, in maniera tale da renderlo facilmente identificabile e collegabile.
Nel nostro caso, l' URI è così strutturato:

– il dominio http://id.sbn.it/ come radice dell'URI;
– l'identificativo univoco del record.

Un esempio di URI di documento è il seguente: http://id.sbn.it/BVE0391786.
Per quanto riguarda i dati SBN, è necessario tener presente la possibilità che un record possa essere cancellato o fuso con un altro. Allo scopo di non vanificare la funzione dell'URI, sarà dunque necessario prevedere un sistema di re-indirizzamento che permetta, nel caso di fusione fra due record, di essere automaticamente reindirizzati al nuovo URI, tramite il vecchio collegamento, mantenendo così il riferimento all'oggetto originario.
La base dati SBN potrà essere collegata facilmente ad altre fonti esterne (ad es. il Nuovo Soggettario di Firenze, la Classificazione Decimale Dewey, l'Anagrafe delle biblioteche italiane, ISBN, etc.). Oltre a questi archivi, saranno aggiunti collegamenti con i *repository* di Geonames, Edit16, VIAF e OPAC SBN (per quanto concerne le schede di autorità), al fine anche di aumentarne la visibilità all'esterno. I legami alle entità dei suddetti *repository* sono stati introdotti in modo puntuale all'interno del prototipo a titolo esemplificativo e sono stati tradotti in triple RDF, memorizzati e consultabili all'interno del *triple store*.

4.3. Conversione dei dati

La fase successiva ha riguardato la conversione dei dati dal formato UNIMARC al formato RDF.
Il subset di record bibliografici, una volta normalizzato e arricchito nei campi dimostrativi per il prototipo, è stato convertito in triple RDF attraverso la mappatura FRBRoo, applicando strumenti e procedure sviluppati *ad hoc*, per essere poi caricato all'interno di un prototipo di *triple store* e sottoposto a test per verificare la coerenza e correttezza della mappatura utilizzata.

4.4. Pubblicazione

La fase successiva della sperimentazione ha riguardato la pubblicazione dei dati convertiti in RDF, che sono stati riversati all'interno di un *triple store* per renderli

accessibili all'utenza finale. Per la pubblicazione degli stessi sono state prese in considerazione piattaforme che, oltre al *repository* RDF e l'*endpoint* di interrogazione SPARQL, avessero a disposizione interfacce aggiuntive per la costruzione di servizi. La scelta è caduta sulla piattaforma *open source* Aduna Sesame che rispetto all'altra esaminata, OpenLink Virtuoso, ha evidenziato un'ottima efficienza nella gestione di database *in-memory*, considerando anche il numero limitato di record processati dal prototipo.
Le triple RDF disponibili nel triple store Sesame sono accessibili mediante:
– consultazione e download delle risorse sotto forma di file RDF/XML;
– *endpoint* SPARQL, mediante il quale è possibile ottenere informazioni a seguito dell'inserimento di una *query*;
– API che permettono la creazione di servizi aggiuntivi per facilitare le *query* sul *triple store*.

A questa prima fase sperimentale ne seguirà un'altra che prevede l'utilizzo di una piattaforma di creazione delle triple, *storage* e la pubblicazione dei risultati del tutto nuova.
L'ICCU infatti collabora al progetto del Polo digitale che interessa 5 Istituti Culturali di Napoli (Società Napoletana di Storia Patria, Pio Monte della Misericordia, Istituto Italiano per gli Studi Storici, Fondazione Biblioteca Benedetto Croce, Cappella del Tesoro di San Gennaro) che ha tra i suoi obiettivi la realizzazione di una piattaforma *open source* che consentirà la gestione dei complessi processi di lavoro finalizzati alla digitalizzazione, conservazione e divulgazione di beni culturali documentali archivistici, bibliografici e museali (dipinti, oggetti d'arte, disegni). La piattaforma prevede la conversione dei metadati bibliografici digitali in LOD utilizzando la mappatura realizzata dall'ICCU.

4.5. Aspetti da definire

Sono state analizzate infine una serie di criticità emerse nel corso delle attività, che riguardano sostanzialmente i dati del catalogo e il modello prescelto del CIDOC CRM/FRBRoo. Di seguito si riassumono le criticità inerenti i dati e le relazioni SBN:
– presenza di dati non normalizzati: si tratta di dati inseriti dai catalogatori in forma testuale e non codificata che inevitabilmente non risultano indicizzabili e di conseguenza di difficile ricerca;
– l'archivio d'autorità Autori presenta un numero limitato di schede di autorità rispetto al numero complessivo di autori presenti nell'Indice SBN. Basti pensare che, a fronte di circa 4 milioni di record autori in Indice sono presenti soltanto 204.000 schede di autorità controllate. Inoltre, in diversi casi, tali schede sono carenti delle informazioni necessarie all'identificazione univoca dell'autore (ad esempio, data di nascita e morte);

– per quanto riguarda gli altri archivi d'autorità (titolo uniforme, luoghi, marche, ecc.) le relative schede non vengono attualmente esportate in OPAC, in quanto non ne è ancora stata definita la modalità di compilazione;
– le entità FRBR non sono tutte rappresentate nell'architettura SBN, che prevede la descrizione dell'entità Manifestazione (*Manifestation*), ancora parzialmente quella relativa all'Opera (*Work*), mentre non sono rappresentate quelle riguardanti l'Espressione (*Expression*) e l'Esemplare (*Item*);
– impossibilità di riferirsi a identificativi persistenti (ad esempio, BID e VID) per la natura stessa della catalogazione partecipata che prevede fusioni e cancellazioni dei record da parte delle biblioteche;
– la mole di dati contenuti nell'archivio SBN (circa 14 milioni di titoli) rende di difficile attuazione la pulizia e la normalizzazione dei dati stessi.

A proposito del modello FRBRoo le criticità riguardano essenzialmente la difficoltà di tradurre la totalità dei dati e dei legami presenti in SBN. Nel caso dei periodici, ad esempio, non è stato possibile completare la mappatura fra SBN e FRBRoo in quanto nel modello non sono rappresentati i dati relativi ad alcuni legami (tag 423; 431; 434; 447, rispettivamente "Pubblicato con", "Continuazione in parte di", "Ha assorbito", "Si fonde con").
Si attende il consolidamento del lavoro del gruppo costituto in ambito ISSN, che ha elaborato l'evoluzione e l'estensione di FRBRoo alle risorse in continuazione[10]. Il documento prodotto, PRESSoo (versione 1.0)[11], non è ancora uno standard IFLA, ma può essere considerato un valido strumento per la mappatura delle relazioni tra risorse in continuazione.
Il Gruppo LOD SBN sta inoltre seguendo a livello internazionale altre iniziative, quale il progetto di sviluppo della rappresentazione di UNIMARC in RDF presentato nel 2013 al convegno IFLA con il documento "The UNIMARC in RDF project: namespaces and linked data"[12].

5. La collaborazione in corso con Wikimedia Italia

Come già accennato nell'introduzione, lo scorso 9 marzo l'ICCU ha sottoscritto un accordo triennale con l'Associazione Wikimedia Italia (WMI), che stabilisce una collaborazione volta al riutilizzo e all'integrazione dei dati e dei materiali dei progetti ICCU con i progetti Wikimedia e con il progetto OpenStreetMap[13].

[10] Il gruppo è costituito da: ISSN, International Centre ISSN IC, ISSN Review Group e Bibliotèque Nationale de France. Si veda anche http://www.issn.org/the-centre-and-the-network/our-partners-and-projects/pressoo/.
[11] Il documento è disponibile al link http://www.issn.org/wp-content/uploads/2014/02/PRESSoo_1-0.pdf.
[12] Il documento è disponibile al link http://library.ifla.org/156/1/222-willer-en.pdf.
[13] L'annuncio ufficiale è disponibile al link: http://www.iccu.sbn.it/opencms/opencms/it/archivionovita/2015/novita_0008.html.

SISTEMA MODUS

La realizzazione pratica dell'accordo è affidata a un "wikipediano in residenza", ossia a un membro della comunità wikimediana che si occupa tanto di operare direttamente sui progetti Wikimedia, quanto di formare il personale dell'Istituto all'uso dei progetti[14].

In questi primi mesi[15], la collaborazione si è concentrata sull'Anagrafe delle Biblioteche e CulturaItalia, che sono stati considerati un buon punto di partenza per valutare la fattibilità di una collaborazione. Al 16 novembre 2015, sono stati generati 111 collegamenti da Wikipedia verso l'Anagrafe delle biblioteche[16] e 306 collegamenti verso CulturaItalia[17].

In aggiunta, si è approfondito il lavoro di sincronizzazione delle schede di autorità di SBN di livello AUF con il progetto Wikidata, database di conoscenza libera e multilingue appartenente alla famiglia dei progetti Wikimedia[18].

La sincronizzazione è iniziata il 6 aprile 2013 su input della comunità di Wikidata, attraverso la creazione della proprietà[19] P396, che identifica univocamente gli "identificatori SBN"[20] e genera un collegamento fra un elemento[21] e la relativa scheda su SBN. Il collegamento è gestito tramite il link di re-indirizzamento apposito http://id.sbn.it/af/ seguito dal codice della scheda[22]. Al 16 novembre 2015, sono già 13.471 i collegamenti confermati fra Wikidata e SBN sulle 64.502 schede di livello AUF finora rese disponibili.

Questo lavoro di sincronizzazione fra il lavoro svolto indipendentemente dall'ICCU e quello svolto dalla comunità wikimediana è ovviamente solo un primo passo, volto a preparare la base per la seconda fase di collaborazione in cui l'ICCU potrà sfruttare (previa validazione) i dati provenienti dai progetti Wikimedia e la comunità wikime-

[14] Il termine di "wikipediano in residenza" è derivato da quello di "artista in residenza", ossia dalla definizione di un artista contemporaneo che viene ospitato (spesso in forma retribuita) da un privato o da una istituzione e che "ricambia" l'ospitalità ricevuta tramite la creazione di opere d'arte che restano poi nella disponibilità del soggetto ospitante.

[15] Per motivi di sintesi, ci si concentrerà qui solo sui principali punti di collaborazione finora sviluppati. Una panoramica più completa dei risultati ottenuti è disponibile al link https://it.wikipedia.org/wiki/Progetto:GLAM/ICCU.

[16] La lista di collegamenti è disponibile al link https://it.wikipedia.org/wiki/Categoria:Codice_ISIL_letto_da_Wikidata.

[17] La lista di collegamenti è disponibile al link https://it.wikipedia.org/wiki/Categoria:Codice_CulturaItalia_letto_da_Wikidata.

[18] Il progetto è visibile all'indirizzo https://www.wikidata.org/.

[19] Su Wikidata, una "proprietà" è l'equivalente di un tag UNIMARC.

[20] La proprietà è visibile all'indirizzo https://www.wikidata.org/Property:P396.

[21] Un "elemento" è equivalente alla scheda di un database. Ciascun elemento riguarda un determinato soggetto del "mondo reale" (come un personaggio reale o fittizio, un oggetto, un concetto, un luogo, un evento, eccetera) e contiene al suo interno tanto i dati fondamentali riguardo quel determinato soggetto quanto i collegamenti ai singoli progetti Wikimedia finora integrati con Wikidata. Al 16 novembre, i progetti integrati sono 617 in più di 290 lingue differenti.

[22] Per esempio, il link http://id.sbn.it/af/IT\ICCU\CFIV\002049 corrisponde alla scheda di autorità in OPAC su Giacomo Leopardi (IT\ICCU\CFIV\002049).

SISTEMA MODUS

diana potrà sfruttare le operazioni di controllo e validazione effettuate dall'ICCU, in una dinamica di interscambio dei dati che sia profittevole per entrambi.

6. Sviluppi futuri

L'Istituto, dopo la necessaria fase di test e la messa a punto della mappatura con i dati di SBN, intende proseguire il percorso della condivisione di altre basi dati in LOD sia gestite al suo interno, sia partecipando ad altri progetti di settore oltre quello con WMI. Nello specifico ha costituito un gruppo di lavoro finalizzato all'analisi dei LOD per il profilo della base dati EDIT16 (Censimento nazionale delle edizioni italiane del XVI secolo) per valorizzare le risorse di tale base dati con l'arricchimento del contenuto informativo e della rete di relazioni.

Nel lavoro di mappatura dei dati con il modello concettuale di riferimento, si partirà dalla documentazione prodotta per i LOD del Catalogo SBN e verrà sviluppata la mappatura degli elementi aggiuntivi e peculiari (marche, luoghi, editori) per la conversione in RDF del profilo dati di EDIT16.

L'ICCU inoltre sta collaborando con il progetto "Polo digitale degli Istituti culturali di Napoli" per il quale fornirà la documentazione prodotta dal gruppo di lavoro e avvierà quindi una fase di sperimentazione che prevede l'integrazione di basi dati bibliografiche archivistiche e museali che verranno pubblicate in LOD.

L'ultima consultazione dei siti Web è avvenuta nel mese di dicembre 2015.

Il progetto Clavius on the Web entra nelle scuole

Andrea Marchetti - *Istituto di informatica e telematica (CNR-Pisa)*
Simone Marchi - *Istituto di linguistica computazionale "Antonio Zampolli" (CNR-Pisa)*
Martín Maria Morales - *Archivio storico della Pontificia Università Gregoriana*

1. Clavius on the Web (CoW)

Il progetto CoW costituisce il frutto della collaborazione tra l'Archivio storico della Pontificia Università Gregoriana (APUG), l'Istituto di informatica e telematica (IIT) e l'Istituto di linguistica computazionale "Antonio Zampolli" (ILC), questi ultimi due afferenti al Consiglio nazionale delle ricerche di Pisa. CoW è iniziato nel 2012, anno in cui ricorreva il quarto centenario della morte del matematico gesuita Christophorus Clavius (1537-1612). Tedesco di Bamberga, professore presso il Collegio Romano, Clavius era considerato una delle massime autorità dei suoi tempi per ciò che concerneva le scienze matematiche. Egli fu, infatti, il maestro dei primi quadri scientifici della Compagnia di Gesù e grazie al suo insegnamento alcuni missionari come Matteo Ricci (1552-1610) o Giovanni Antonio Rubino (1578-1643) riuscirono a condurre con mezzi non esclusivamente teologici la loro opera di missione in Asia. Oggi si ricorda Clavius principalmente per il suo contributo alla riforma gregoriana del calendario, approvata nel 1582 su un modello proposto da Luigi Lilio (1510-1576), successivamente perfezionato dal gesuita tedesco[1].

La ricorrenza del centenario è stata l'occasione per organizzare il convegno "Christophorus Clavius (1538-1612) alle soglie della scienza. Il suo magistero, le sue reti" (Roma, Pontificia Università Gregoriana, 19 ottobre 2012), evento nel quale è stata presentata l'idea progettuale di CoW.

1.1. Obiettivi

L'obiettivo iniziale del progetto è stato quello di valorizzare il fondo dei manoscritti riguardanti Clavius conservati presso l'Archivio storico della Pontificia Università Gregoriana, in particolare i due codici contenenti la maggior parte della corrispondenza del matematico gesuita giunta sino a noi[2]. Prima di procedere a qualunque

[1] *Gregorian reform of the calendar. Proceedings of the Vatican conference to commemorate its 400th anniversary, 1582-1982,* edited by George V. Coyne, Michael A. Hoskin and Olaf Pedersen, Città del Vaticano: Specola Vaticana, 1983.

[2] Si tratta dei codici APUG 529 e 530: le lettere sono per la maggior parte indirizzate a Clavius. Questa corrispondenza è stata studiata e pubblicata in: Christophorus Clavius, *Corrispondenza,* edizione critica a cura di Ugo Baldini e Pier Daniele Napolitani, Pisa: Università di Pisa, Dipartimento di matematica, 1992.

altra operazione, i due codici sono stati interamente restaurati, nella convinzione che qualunque iniziativa di digitalizzazione debba in primo luogo preoccuparsi della conservazione dei documenti originali usati nel progetto.
Il restauro ha previsto la sfascicolazione dei due codici per ricondurre le 300 lettere ivi contenute al loro formato originario. Tale operazione permette oggi di consultare le lettere singolarmente - garantendone una più adeguata conservazione - e ne ha inoltre consentito una migliore digitalizzazione.
L'idea progettuale si è basata sulla realizzazione di una piattaforma integrata che consentisse la navigazione fra vari livelli di analisi. Le lettere sono state infatti analizzate da un punto di vista linguistico, semantico e lessicale attraverso l'uso di strumenti specifici.
In particolare l'analisi linguistica - preceduta dalla codifica in TEI/XML dei testi e condotta finora solo su quelli in latino - è stata realizzata con un tool automatico, denominato *ClaviusLemmata*[3]. La sfida principale lanciata da questo tipo di analisi è consistita nell'assenza di adeguate basi di dati linguistiche per il latino umanistico. Come noto, infatti, le più importanti risorse per il latino sono basate su quello classico[4], motivo per cui il loro uso per la corrispondenza di Clavius - peraltro densa di terminologia scientifica - ha dato risultati che in alcuni casi è stato necessario correggere manualmente attraverso un lavoro di *proofreading*.
L'analisi semantica è stata condotta usando il tool di annotazione BRAT[5], opportunamente configurato al fine di annotare le entità ritenute significative citate nei testi, come persone, luoghi, eventi, opere.
Infine, per quanto concerne l'analisi lessicale, è stato creato un lessico computazionale dove sono state strutturate e formalizzate le definizioni dei termini matematici e astronomici date da Clavius nell'edizione in cinque volumi in folio dell'*Opera Mathematica*, pubblicata a Magonza tra il 1611 e il 1612[6].
I dati generati da queste analisi sono confluiti in un'unica base di dati successivamente esposta in forma di linked data, che costituisce il *Clavius linked dataset*.
Il principale problema che ci si è trovati ad affrontare nel lavoro di analisi ha riguardato l'integrazione delle annotazioni, visto che spesso uno stesso termine risulta essere annotato da diversi punti di vista: ad esempio "Clavius" può essere considerato sia dal punto di vista grammaticale, quindi un nome proprio di persona, sia come personaggio citato. L'integrazione è stata dunque assicurata dall'uso

[3] L'analisi si articola in quattro fasi: 1) separazione delle frasi; 2) tokenizzazione; 3) assegnazione della categoria grammaticale e dei tratti morfosintattici; 4) lemmatizzazione.
[4] La *treebank* utilizzata è stata la Latin Dependency Treebank (LTD), disponibile all'indirizzo <http://nlp.perseus.tufts.edu/syntax/treebank/>. David Bamman - Gregory Crane, *The latin dependency treebank in a cultural heritage digital library*, in: *Proceedings of the workshop on language technology for cultural heritage data (LaTeCH 2007)*, Prague: Association for Computational Lingusitics, 2007, p. 33-40.
[5] http://brat.nlplab.org/.
[6] Ad oggi il lessico conta 158 termini.

SISTEMA MODUS

di un sistema di identificazione univoco delle parole all'interno dei testi, il Canonical Text Service (CTS). Il CTS permette di associare a ogni parola o gruppo di parole consecutive un Uniform Resource Name (URN).
Un altro degli obiettivi iniziali del progetto è stato quello di prendere parte al progressivo ingresso delle tecnologie digitali all'interno delle scuole italiane, proponendo idonei strumenti da usare nella didattica, come ad esempio lo *storytelling*. Su questo aspetto si tornerà in 2.

1.2. Il portale Clavius on the Web

Oggi il portale CoW consente di navigare fra diverse tipologie di risorse: riproduzioni dei manoscritti, visualizzazioni delle loro trascrizioni e dei vari livelli di analisi, riproduzioni interattive di alcuni strumenti scientifici in carta, *storytelling*.
Allo scopo di invogliare l'utente a non fermarsi solo sulla pagina iniziale, si è deciso di posizionare già nella Home (Fig. 1) un'anteprima di alcuni dei contenuti disponibili, con l'idea che la visione di una porzione di essi, seppur minima, potesse spingere ad approfondire la ricerca per visualizzare tutte le informazioni e le funzioni associate a quel contenuto.

Figura 1: *La homepage del portale Clavius on the Web*

1.2.1. Lettere e strumenti

Sebbene la digitalizzazione delle lettere presenti nei due codici APUG sia stata completata, al momento sul portale è possibile accedere solo a una parte di esse, dato il complesso e impegnativo lavoro di analisi cui vengono sottoposte.
Nella Home, come già accennato, sono visibili piccole anteprime delle lettere. Cliccando su di esse si accede a una pagina dove viene presentata la riproduzione del manoscritto cui è affiancata la trascrizione del testo: lo scorrimento dell'imma-

Figura 2: *La barra di accesso alle funzionalità disponibili*

gine è allineato a quello della trascrizione. Grazie alla codifica TEI delle lettere è stato possibile realizzare una corrispondenza tra porzioni del manoscritto e trascrizione: passando il puntatore su una porzione di testo trascritta, automaticamente viene evidenziata la parte di immagine corrispondente e viceversa[7].

Dalla Home è inoltre possibile accedere ad altre funzionalità corrispondenti ai livelli di analisi cui sono state sottoposte le lettere. Passando il cursore sull'immagine compare infatti un piccola barra (Fig. 2) da cui si accede, nell'ordine, a: un editor TEI; un tool di analisi linguistica; la visualizzazione parallela di testo e immagine; la visualizzazione di tutte le annotazioni presenti nella lettera; il grafo a nodi delle entità annotate nel testo.

I primi due strumenti sono stati appositamente sviluppati per il lavoro con le scuole, di cui si parlerà nel prossimo paragrafo (2); della visualizzazione parallela testo e immagine si è già detto poco sopra.

La visualizzazione in un'unica risorsa sia del testo sia dei risultati dell'analisi linguistica e semantica è stata attuata allo scopo di facilitare la lettura e la traduzione delle lettere scritte in latino, lingua di non facile accesso per tutti gli utenti (Fig. 3). Sotto ogni unità linguistica viene indicata la categoria di appartenenza - verbo, sostantivo, aggettivo e così via -, distinta attraverso l'uso di differenti simboli gra-

[7] Al momento questa funzione richiama la codifica TEI del testo, in particolare il tag <lb>.

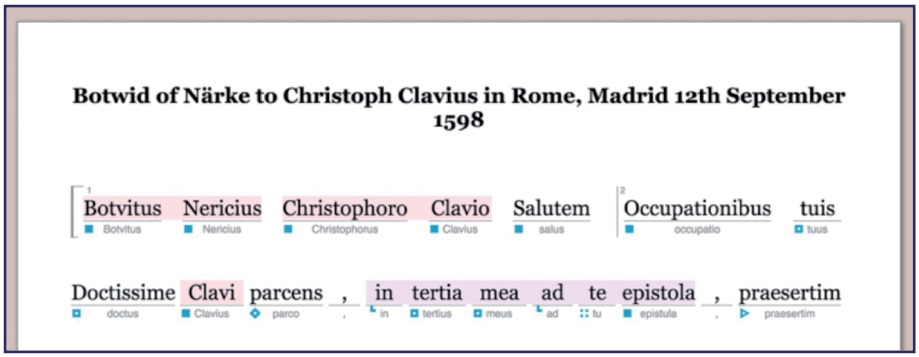


Figura 3: *Visualizzazione di tutte le annotazioni*

fici, i quali, una volta memorizzati dall'utente, risultano più facilmente riconoscibili rispetto ad altro genere di indicazioni[8]. Accanto a ciascun simbolo è indicato inoltre il lemma della parola presente nel testo. Non solo: accanto alle informazioni linguistiche vengono visualizzate anche le annotazioni semantiche, per cui tutte le entità citate nel testo come persone, luoghi o date, sono evidenziate con colori diversi a seconda della classe di appartenenza. Quest'ultima viene visualizzata passando con il cursore sulla porzione di testo interessata dall'annotazione.

Infine, il grafo a nodi contiene sia le entità citate nelle lettere, sia i metadati delle stesse. L'accesso avviene attraverso il servizio LodLive[9], grazie al quale è possibile navigare nella base di conoscenza con una modalità simile all'attuale struttura dei dati presenti sul Web.

Nella Home è presente anche una sezione *Instruments,* dove è stata inserita la digitalizzazione di alcune volvelle, strumenti in carta ritrovati nella corrispondenza di Clavius composti da dischi di carta sovrapposti, movibili grazie ad un perno centrale cui sono ancorate (Fig. 4). Alcune tra queste sono da collocarsi nello scambio di lettere tra il gesuita tedesco e il suo compagno e allievo Jean Deckers (1550-1619), dedicate ad alcuni aspetti della cronologia cristiana. Queste volvelle servirono a dimostrare alcune posizioni sostenute da Deckers in merito alla cronologia della vita di Cristo. Solo grazie all'animazione digitale è stato possibile capire quale fosse il loro funzionamento: la rotazione simulata dei dischi di carta che le compongono ha infatti permesso di evitare di compiere la stessa operazione sugli originali, fatto che ne avrebbe probabilmente compromesso l'integrità.

[8] Il "minilinguaggio" simbolico è stato pensato per facilitare la memorizzazione dei simboli e la loro associazione alla parte del discorso corrispondente. Ad esempio, un quadrato pieno indica un sostantivo e uno vuoto un aggettivo, mentre un triangolo pieno indica un verbo e uno vuoto un avverbio. Altri segni più piccoli sono usati per preposizioni e congiunzioni, mentre varianti dei simboli discussi precedentemente sono usate per pronomi (quattro piccoli quadrati, disposti in modo da somigliare al segno quadrato del sostantivo) e participi (quadrato vuoto come l'aggettivo, ma ruotato in modo da somigliare più al simbolo triangolare del verbo).

[9] http://lodlive.it/.

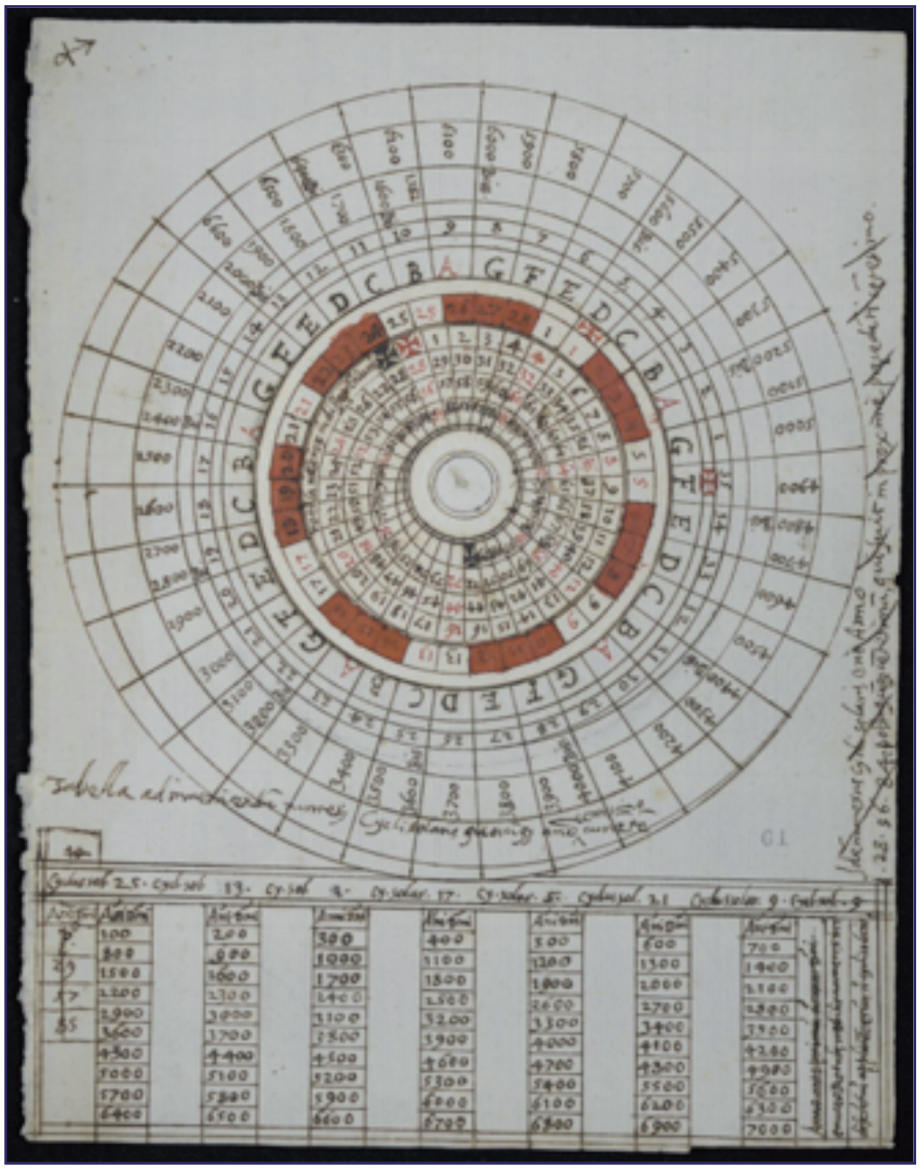


Figura 4: *Volvella di Jean Deckers (APUG 530, c. 61r)*

1.2.2. Storie

Nella sezione *Storie* sono presenti alcuni esempi di *storytelling* riguardanti tematiche inerenti Clavius. Una delle storie è stata dedicata alla vita del gesuita tedesco: in questa è possibile "navigare" nella sua biografia attraverso una Timemap - una mappa abbinata a una linea temporale - dove agli eventi sono legate sia informazioni relative ai luoghi in cui avvennero, sia brevi descrizioni degli stessi.

Un'altra storia è dedicata alla riforma gregoriana del calendario, cui Clavius prese parte come membro della commissione pontificia appositamente istituita. Questo *storytelling* è stato realizzato durante il workshop *Clavius@Work* (vedi **2**), grazie alla collaborazione del prof. Pardo Fornaciari, il quale - vista la pluriennale esperienza come educatore nei licei - ha proposto un approccio che potesse stimolare il più possibile la curiosità dei ragazzi. Si è dunque deciso di seguire in apertura un metodo simile a quello dei *mockumentary*, giocando sulle conseguenze della riforma del calendario avutesi nel Cinquecento - in particolare la soppressione dei dieci giorni dal 5 al 14 ottobre 1582 - proiettate nella società di oggi.

2. Clavius@School

Durante il 2014 sono stati organizzati due workshop all'interno di CoW: il primo dal titolo *Clavius@Work* - svoltosi presso l'Area della Ricerca del CNR di Pisa nei giorni 11-13 giugno 2014 - era destinato a completare alcune attività relative al progetto con l'ausilio di tecnici ed esperti provenienti da altri istituti. Il secondo, organizzato invece a Roma presso la Pontificia Università Gregoriana nei giorni 12-14 novembre 2014, ha avuto come obiettivo primario quello di coinvolgere il mondo della scuola all'interno del progetto CoW. Seguendo la denominazione del primo workshop, questo secondo evento è stato intitolato *Clavius@School* (Fig. 5).

L'intenzione di entrare nel mondo della scuola, come accennato, era già presente fra gli obiettivi iniziali del progetto, sebbene le sue modalità di attuazione non fossero ancora state stabilite. L'idea era nata con lo scopo di coinvolgere principalmente i professori, i quali avrebbero potuto trarre giovamento dall'uso della piattaforma CoW all'interno delle classi, dunque arricchendo l'esperienza formativa degli studenti attraverso l'utilizzo di un supporto innovativo da affiancare alla didattica frontale. In questa prospettiva era stato invitato un professore già al workshop *Clavius@Work*, grazie al quale si era potuto creare uno *storytelling* adatto a un pubblico di studenti di scuola superiore (vedi **1.2.2**).

Col passare del tempo si è pensato però che il coinvolgimento diretto degli studenti avrebbe potuto essere più fruttuoso, a maggior ragione se messo in opera dal team di ricerca che aveva lavorato al progetto. Un primo tentativo orientato in questa direzione è stato messo in atto durante l'Internet Festival 2014 - svoltosi a Pisa tra il 9 e il 12 ottobre - dove alcuni studenti universitari sono stati invitati a usare gli strumenti di analisi sviluppati in CoW.

L'organizzazione di *Clavius@School* è stata strutturata mettendo a frutto l'esperienza maturata durante gli eventi sopra menzionati avendo, in questo caso, gli studenti di scuola superiore come utenti di riferimento.

Figura 5: *Locandina del workshop Clavius@School*

2.1. Il lavoro degli studenti

Grazie alla disponibilità e alla collaborazione del Liceo Ennio Quirino Visconti di Roma[10] è stato possibile coinvolgere a *Clavius@School* un'intera classe e alcuni studenti provenienti da altre classi, per un totale di 24 partecipanti.

Le due giornate di lavoro sono state suddivise in quattro sessioni, di cui tre di lavoro vero e proprio e una quarta destinata al riepilogo di quanto svolto, allo scopo di presentarne i risultati nella sessione plenaria del terzo giorno, quando tutti i

[10] In questa sede ci preme ringraziare in particolare le professoresse Romana Bogliaccino e Paola Vasconi e il professor Giovambattista Galbo; inoltre, questa esperienza non sarebbe stata possibile senza la collaborazione della preside Clara Rech.

SISTEMA MODUS

gruppi hanno illustrato le attività effettuate. Le tre sessioni di lavoro sono state inoltre precedute da una breve introduzione al progetto.
L'obiettivo del lavoro è stato quello di coinvolgere gli studenti nell'esperienza di CoW attraverso tre tipologie di attività:

- trascrizione delle lettere;
- analisi linguistica del latino e *proofreading*;
- analisi semantica e lessicale delle lettere.

Per svolgere queste attività gli studenti sono stati suddivisi in otto gruppi di tre e messi in competizione tra loro: i migliori di ognuna delle tre sessioni hanno ricevuto un premio[11]. I ragazzi si sono confrontati esclusivamente con i coordinatori del gruppo, senza la presenza dei loro insegnanti.
I lavori di trascrizione e analisi sono stati svolti utilizzando i tool sviluppati nel progetto, alcuni appositamente predisposti per essere usati dagli studenti, come i sistemi per la trascrizione dei testi e per il *proofreading*. In particolare per la trascrizione, oltre alle riproduzioni digitali, gli studenti hanno potuto usare anche delle riproduzioni cartacee delle lettere, in modo da assimilare il più possibile la loro attività a quella dei ricercatori che studiano i documenti originali. Il tool di trascrizione doveva consentire inoltre una codifica TEI minima - necessaria poi per l'analisi linguistica automatica del latino - e semplificata, per cui si è provveduto a sostituire ai tag segni di più immediata comprensione[12]: in tal senso questo tool può essere definito un editor TEI a tutti gli effetti.
Alla fine delle attività il dato più interessante è stato scoprire il giudizio degli studenti sul lavoro svolto. La parte giudicata maggiormente affascinante è stata quella della trascrizione delle lettere, dove gli studenti si sono sentiti dei veri e propri ricercatori impegnati nella lettura di documenti antichi, seppure questa sia avvenuta solo attraverso delle riproduzioni degli originali. Il fascino esercitato dai manoscritti è stato tale da suscitare negli studenti il desiderio - prontamente esaudito - di visitare l'archivio allo scopo di vedere da vicino i documenti originali.
Il feedback sulle altre attività è stato anch'esso positivo: l'analisi automatica del latino è stata considerata utile per la traduzione e la comprensione dei testi; il *proofreading* è stato giudicato un utile esercizio per migliorare le competenze grammaticali latine; l'annotazione semantica e lessicale hanno invece creato alcune difficoltà, implicando un maggiore sforzo di esegesi del testo, connesso a una conoscenza non superficiale dell'argomento trattato. Tuttavia, gli studenti si sono dedicati con entusiasmo alla costruzione di una ontologia dei concetti matematici, cercando di associare ai principali concetti la definizione latina corretta, tra quelle fornite alla rinfusa dal coordinatore dell'attività.

[11] I premi sono consistiti in materiale relativo al progetto come segnalibri o riproduzioni di lettere, libri e gadget tecnologici.
[12] Ad esempio il tag <lb> è stato reso con il simbolo "$".

Gli studenti hanno inoltre proposto miglioramenti da apportare al sistema riguardanti questioni legate all'usabilità e all'accesibilità (grafica, fruibilità) fino all'aggiunta di nuove funzionalità (traduzione delle lettere, creazione di mappe concettuali).
L'esperienza è stata in generale molto positiva: il progetto ne ha beneficiato grazie ai preziosi feedback ricevuti in merito all'uso e alle funzionalità della piattaforma; gli studenti hanno sperimentato una modalità inconsueta di apprendimento, strettamente connessa all'uso delle tecnologie digitali e a diretto contatto con gli sviluppatori dei sistemi, esperienza che li ha resi consapevoli della complessità delle operazioni necessarie a rendere i documenti ricercabili e fruibili online.

2.2. Gli altri gruppi di lavoro

Parallelamente a quello degli studenti, hanno lavorato altri due gruppi. Il gruppo "Professori" ha visto la partecipazione degli insegnanti responsabili degli studenti coinvolti in *Clavius@School* e ha avuto come obiettivo principale quello di immaginare le funzionalità che dovrebbe avere una piattaforma dedicata alla didattica. Partendo dall'individuazione di possibili percorsi interdisciplinari nella corrispondenza di Clavius, si è proceduto a delineare una serie di attività che gli studenti potrebbero svolgere con il supporto di una piattaforma CoW dedicata alla didattica, il cui obiettivo finale dovrebbe essere la pubblicazione di uno storytelling su un argomento di ricerca assegnato dai docenti.
Il gruppo "Esperti" ha infine visto la partecipazione di specialisti provenienti da vari campi di specializzazione: catalogatori, filologi, filosofi, sociologi, informatici, ingegneri, restauratori, archivisti. Ognuno ha messo a disposizione il proprio bagaglio di competenze allo scopo di discutere delle problematiche relative all'uso delle tecnologie nei beni culturali, comprendendo temi quali il restauro digitale dei manoscritti, la conservazione dei documenti digitali, la metadazione e l'organizzazione della conoscenza in rete.

3. Clavius on the Web 2

3.1. Obiettivi

Negli anni 2015 e 2016 CoW continuerà il suo percorso concentrandosi sul consolidamento di quanto già realizzato dal punto di vista tecnico e al tempo stesso aumentando la quantità dei contenuti, tentando di arrivare a rendere disponibile online l'intera opera di Clavius conservata presso l'Archivio storico della Pontificia Università Gregoriana.
Il lavoro svolto con le scuole prenderà inoltre una forma più definita - integrandosi maggiormente con la didattica dell'anno in corso - con l'idea di condurre ulteriori sperimentazioni con altre classi. Al Liceo Visconti, che ha confermato la sua disponibilità a continuare questo progetto, si è unita anche un'altra scuola superiore di

Roma, il Liceo Virgilio - che oltre all'indirizzo classico ha anche quelli scientifico e linguistico. La partecipazione di altri istituti - anche di diversa tipologia, come istituti tecnici - è comunque aperta e benvenuta. A tale scopo è stato allestito un forum sul portale CoW dove è possibile trovare e chiedere informazioni al riguardo[13].
La collaborazione con il mondo della scuola porterà inoltre ad implementare nell'architettura del sistema le funzionalità per la didattica immaginate in *Clavius@School* (cfr. 2.2), grazie alla consulenza fornita dagli insegnanti.

3.2. Epimeleia

Il contatto diretto con il mondo della scuola ha consentito di iniziare una riflessione sull'uso responsabile delle tecnologie - in particolare di quelle digitali - e sul loro impatto nella società contemporanea. La scuola è infatti il luogo dove elaborare una posizione allo scopo di proteggersi dalle minacce di quello che il filosofo francese Bernard Stiegler ha identificato come "populismo industriale". Questa estensione del populismo genera una proletarizzazione degli utenti, confinandoli a essere solo i consumatori dei prodotti di quell'industria e relegando loro al "regno dell'ignoranza"[14]. Il progetto CoW pretende di rovesciare nell'ambito scolastico questa logica - per quanto limitatamente ad alcune decine di studenti di liceo - e di implementarne delle altre di tipo associativo e partecipativo.
Nella stessa prospettiva è stata inaugurata anche una serie di incontri dal titolo "Epimeleia"[15], parola greca che indica la cura di sé[16], intesa in questo caso come la presa di coscienza della natura farmacologica della tecnologia, allo stesso tempo rimedio e veleno - dal greco *pharmakon* - per la vita degli esseri umani. Visto il rapporto sempre più totalizzante che la società contemporanea ha instaurato con le tecnologie digitali, esse hanno assunto un ruolo determinante nella formazione della coscienza degli individui. Rendere consapevoli le giovani generazioni di questa insidiosa ambivalenza sarà uno degli obiettivi del progetto CoW nei prossimi due anni.

[13] http://claviusontheWeb.it/forum/.
[14] Bernard Stiegler, *Réenchanter le monde*, Paris: Flammarion, 2006.
[15] Il primo incontro si è svolto a Roma il 23 aprile 2015 presso la Pontificia Università Gregoriana; il calendario dei successivi eventi sarà pubblicato sul sito del progetto.
[16] Bernard Stiegler, *Prendersi cura. Della gioventù e delle generazioni*, Napoli-Salerno: Orthotes, 2014.

L'ultima consultazione dei siti Web è avvenuta nel mese di dicembre 2015.

I possessori nella Biblioteca Nazionale di Napoli: un Archivio di immagini

Simona Pignalosa
Biblioteca Nazionale di Napoli

Il trattamento delle provenienze librarie della Biblioteca Nazionale di Napoli (BNN) avviene in due distinti contesti: uno è il catalogo possessori e provenienze nel Polo di Napoli SBN (Servizio bibliotecario nazionale), l'altro, oggetto di quest'intervento, è l'Archivio Possessori, un data base digitale, fruibile a partire dal sito Web della biblioteca, che contiene le immagini di segni di possesso usati da precedenti proprietari: timbri, legature, ex libris, annotazioni manoscritte. L'impegno a incrementare entrambi gli ambienti, che ovviamente interagiscono strettamente, proviene dalla consapevolezza dell'opportunità che lo studio delle provenienze offre alla ricostruzione della storia, dei percorsi e degli intrecci di antiche biblioteche. Opportunità tanto più stimolante se applicata a un organismo complesso, stratificato e straordinariamente ricco come la BNN.

Possessori in SBN

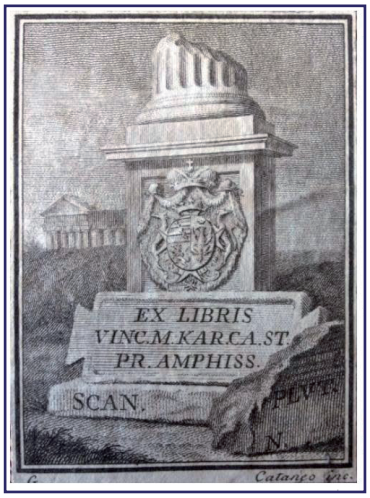
Dopo un periodo di sperimentazione, la gestione delle provenienze della BNN in SBN ha preso avvio in modo sistematico nel 2004, come arricchimento della schedatura dei fondi storici. In SBN sono catalogati gli antichi proprietari dei libri, ciascuno dei quali, persona o ente, è formalizzato secondo le regole di normalizzazione degli autori e corredato, ove possibile, da riferimenti repertoriali e note biografiche e dai necessari rinvii. Ogni possessore è legato a tutti gli inventari di documenti che presentano gli indizi utili a determinarne quella precedente appartenenza. Di conseguenza, il numero di inventari legati a un dato possessore consente di ricostruire virtualmente la provenienza di interi nuclei librari, attualmente fusi nei fondi antichi della BNN. Le note di esemplare relative a ciascuna unità bibliografica registrano fedelmente la presenza di segni di possesso, descrivendoli e precisandone la posizione nel documento.
L'archivio SBN è condiviso dalle 203 biblioteche del Polo NAP, anche se non tutte lo incrementano. Attualmente[1] le notizie-possessori per la BNN sono circa 900, su un totale di 1550 relative al Polo nel suo complesso.

[1] I dati si riferiscono alla situazione del giugno 2015.

SISTEMA MODUS

L'Archivio Possessori

L'Archivio Possessori[2] è uno strumento diverso rispetto a SBN, perché non catalo-ga i precedenti proprietari dei libri, ma i contrassegni che essi hanno apposto ai volumi schedati: ex libris, legature, timbri o note manoscritte. Di ciascun segno di possesso si offre l'immagine digitalizzata, corredata da una scheda descrittiva (Fig. 1-2).



La genesi

L'Archivio nasce come strumento accessorio alla catalogazione in SBN, ma assume presto una sua autonoma utilità. L'inizio del tratta-mento sistematico delle provenienze in SBN è coinciso con importanti progetti di recupero del retrospettivo, culminati con la partecipa-zione della BNN al progetto Google[3]. Gruppi di catalogatori esterni si sono succeduti ad af-fiancare il personale interno nella schedatura dei fondi antichi della BNN. Le collezioni inte-ressate dai progetti sono tra le più significati-ve della biblioteca, a partire dalla cosiddetta Sala Farnese, il cui nome evoca la collezione

Figura 1 (in alto): *Carafa, ex libris;* Figura 2: *Carafa, scheda*

[2] http://www.bnnonline.it/index.php?it/330/archivio-possessori.

[3] Il progetto ha comportato la digitalizzazione di collezioni storiche in grandi biblioteche, preceduta dal necessario recupero catalografico.

SISTEMA MODUS

che Carlo di Borbone ereditò dalla madre, Elisabetta Farnese, e che portò con sé a Napoli nel 1734.

La presenza dei catalogatori esterni ha reso indifferibile la necessità di approntare uno strumento che riproducesse i segni di possesso che venivano individuati e identificati via via che procedevano le operazioni di schedatura. Inizialmente sono stati digitalizzati gli ex libris, i timbri, le note manoscritte, gli stemmi, conservando le copie cartacee a disposizione dei catalogatori. L'evoluzione di questo primitivo stadio ha portato a concepire un archivio digitale di immagini da inserire nell'Intranet della Biblioteca, pensato per essere sfogliato dai bibliotecari impegnati nel riconoscimento di un dato segno di possesso. Presto è parsa naturale la scelta successiva: diffondere i risultati della nostra ricerca e rendere fruibile l'Archivio nel Web. Nel marzo 2014 è stato pubblicato l'Archivio Possessori, che offre alla comunità degli studiosi quello che inizialmente era stato pensato come uno strumento ad uso strettamente interno: un archivio di immagini, corredate da schede organizzate in campi interrogabili.

La struttura finale del data base, ospitato nel sito della BNN[4], è stata progettata grazie alla collaborazione tra l'Ufficio Libro Antico e i Servizi Web e di Biblioteca Digitale della biblioteca, che ne hanno curato la realizzazione, attraverso il creatore di moduli del CMS (Content Management System) Museo & Web con cui è realizzato il sito Web della Biblioteca Nazionale di Napoli[5] .

Figura 3: *Introduzione*

[4] http://www.bnnonline.it/

[5] Il CMS è stato prodotto dal Gruppo Meta nell'ambito del progetto Minerva Europe su iniziativa del MiBACT- Ministero per i beni, le attività culturali e il turismo, ed è sviluppato con PHP 5 e MySql.

La sintetica scheda che accompagna ciascuna immagine è strutturata in campi, tutti interrogabili.
La ricerca dei segni di possesso può quindi essere effettuata da diversi accessi:

- si può inserire il nome presunto del possessore nel campo **Possessore**, dove si visualizzerà la/le scheda/e intestate a quel nome, o il rinvio ad altra forma accettata. La ricerca è possibile anche per singola parola del nome. La forma del nome scelta è quella più comune, mentre la forma standard, utilizzata nell'archivio Possessori del Polo di Napoli di SBN, è riportata nel campo **Forma normalizzata**, e il relativo numero identificativo nel campo **PID**. Gli enti religiosi sono schedati sotto il nome, omettendo qualificazioni come: convento, chiesa, monastero ... (per esempio. 'San Domenico Maggiore' in luogo di 'Convento di San Domenico Maggiore')
- se si desidera visualizzare una determinata categoria di segni di possesso (ex libris, legature ecc...), selezionare l'apposita voce dal menù a tendina del campo **Tipologia**. Nello stesso menù è possibile selezionare i timbri storici della BNN. E' inoltre possibile visualizzare i contrassegni i cui possessori non sono stati identificati, nell'ipotesi di un eventuale riconoscimento da parte degli utenti
- se si desidera individuare un segno di possesso a partire da scritte presenti in esso, come motti o formulazioni complesse di possesso, digitare tali scritte, o parti di esse, nel campo **Descrizione**, dove esse sono trascritte tra virgolette. Non si sono riportate le scritte che recano il solo nome del possessore accompagnato da formule come 'ex libris' o simili.
- il campo **Descrizione** contiene anche eventuali brevi note biografiche sul possessore, note sul segno di possesso, con eventuali riferimenti repertoriali, nonché i rinvii da altre forme del nome del possessore esistenti nell'archivio. Il repertorio più spesso citato è: Egisto Bragaglia, *Gli ex libris italiani dalle origini alla fine dell'Ottocento*, Milano, 1993
- il campo **Collocazione** fornisce l'indicazione del volume da cui si è tratta l'immagine schedata

Figura 4: *Descrizione della scheda*

La ricerca

L'Archivio Possessori è accessibile dalla pagina iniziale nel sito della BNN, oppure selezionando *Cataloghi speciali*, o ancora attraverso la pagina dell'*Ufficio Libro Antico*. La ricerca vera e propria è preceduta da una breve nota introduttiva con la descrizione della scheda e delle più efficaci strategie di interrogazione (Fig. 3-4), che non escludono la possibilità di una libera esplorazione tra gli ex libris, i timbri, le legature riprodotti.

Figura 5: *Campi di ricerca*

In ciascuno dei campi (Fig. 5), la ricerca può avvenire anche inserendo i dati in forma parziale. Il primo campo, denominato *Possessore*, costituisce l'intestazione della scheda: il nome dell'ente, famiglia o persona che ha apposto un proprio segno di possesso su volumi della BNN è trascritto nella forma più amichevole, ritenuta intuitiva anche per un pubblico di non specialisti. Tale forma non sempre coincide con quella normalizzata usata per il trattamento dei possessori del Polo SBN di Napoli, per il quale ci si è doverosamente attenuti alle indicazioni dell'Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche (ICCU), utilizzando, se esistente, la forma usata nell'archivio autori dell'indice SBN. Le formulazioni più "agili" dell'Archivio Possessori sono pri-

ve degli asterischi che in SBN sono prescritti per gli enti. Gli enti religiosi, normalizzati in SBN Possessori a partire dal tipo di istituzione – chiesa, monastero, convento - sono qui indicati con il solo nome dell'ente (Fig. 6).

Figura 6: *San Gennaro e Santa Teresa di Torre del Greco*

Alla forma usata in SBN è riservato un apposito spazio, *Forma normalizzata*, mentre il campo *PID* riporta il relativo numero identificativo. Com'è ovvio, queste due aree, e in particolare il *PID*, servono prevalentemente come supporto ai catalogatori del Polo SBN di Napoli.
Il campo *Tipologia* consente di selezionare, attraverso un menu a tendina, un determinato insieme di segni di possesso: ex libris, legature, note manoscritte, timbri di persone o di enti religiosi o laici (Fig. 7).

Figura 7: *Tipologie dei segni di possesso*

In questo campo è possibile anche scegliere i timbri storici della biblioteca. In SBN la presenza di questi contrassegni sui volumi non dà luogo ad un legame-possessore, in quanto non indica un proprietario precedente alla BNN. Viceversa, è sembrato opportuno pubblicare nell'Archivio di immagini questi marchi di proprietà della stessa BNN nelle sue varie, successive denominazioni (Reale Biblioteca, Reale Biblioteca Borbonica, Biblioteca Nazionale) proprio perché particolarmente significative per la sua storia (Fig. 8).

Figura 8: *Alcuni timbri storici della BNN*

La voce *Possessori non identificati* rimanda a quei segni usati da precedenti proprietari non individuati, inseriti nell'ipotesi di un riconoscimento da parte degli utenti. Si è ritenuto infatti di utilizzare la flessibilità propria del Web per proporre il nostro Archivio come uno strumento aperto alla collaborazione del pubblico (Fig. 9).

La *Collocazione* riportata è quella del volume dal quale è tratta l'immagine riprodotta.

Nel campo *Descrizione* trovano spazio diversi elementi. Il primo dato è una sintetica nota biografica, con le notizie essenziali della vita dell'antico proprietario, volutamente limitate, nella maggior parte dei casi, all'indicazione dell'attività principale, con luogo e date di nascita e morte. Le fonti utilizzate per reperire tali informazioni spaziano dai principali repertori biografici, bibliografici ed enciclopedici agli *Authority file* italiani e internazionali, dai censimenti di ex libris e timbri ai dati desunti dalle stesse indicazioni di possesso descritte e dai libri su cui sono apposte. Nella stessa area si trovano i riferimenti repertoriali e altre forme del nome dalle quali si sono effettuati rinvii (Fig. 10).

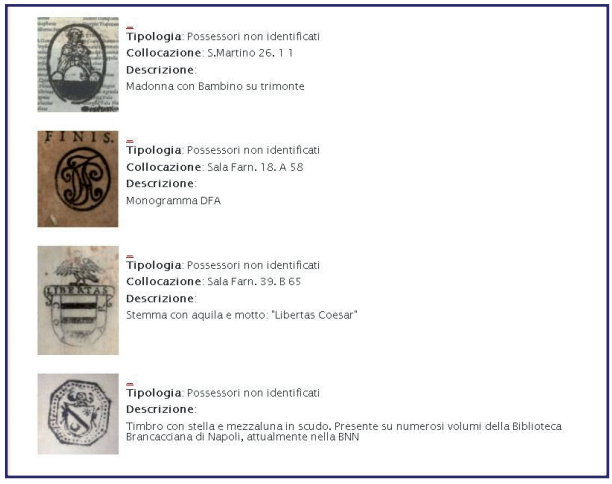


Figura 9: *Possessori non identificati*

Figura 10: *Marefoschi*

Sempre in *Descrizione*, sono riportate, tra virgolette, eventuali scritte presenti nel segno di possesso. Si è evitato di trascrivere formulazioni che indichino il nome del proprietario in forma esplicita o intuitivamente individuabile, mentre sono stati ricopiati motti, sigle, acronimi, citazioni e altre espressioni significative. Queste note costituiscono un ulteriore, importante canale di ricerca per chi si trovi impegnato a decrittare un determinato segno di possesso di cui ignori il titolare. Infatti, il nostro data base è punto di arrivo anche di una interrogazione per parola chiave effettuata in un qualsiasi motore di ricerca (Fig. 11-13).

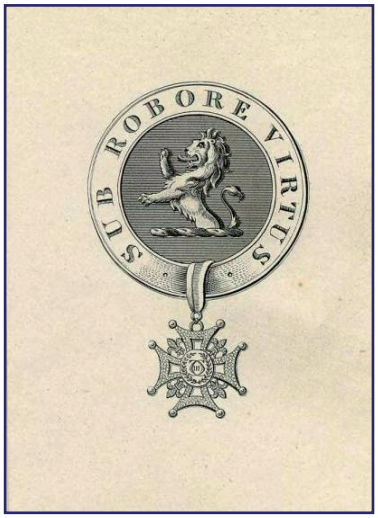


Figura 11: *Un ex libris da identificare*

Accessibilità dai motori di ricerca

Le schermate riprodotte di fianco e nella pagina successiva illustrano molto bene uno dei maggiori pregi dell'Archivio Possessori, la sua interrogabilità direttamente dalla rete. Infatti, se una ricognizione all'interno dell'Archivio offre tutte le opportunità che sono state descritte, una parola o frase lanciata in un qualsiasi motore di ricerca rimanda al nostro Archivio. Questa caratteristica, unita all'altra particolarità del data base, l'offerta delle immagini di tutti i segni censiti, costituisce un'importante facilitazione nello studio della materia. È agevolmente intuibile la conve-

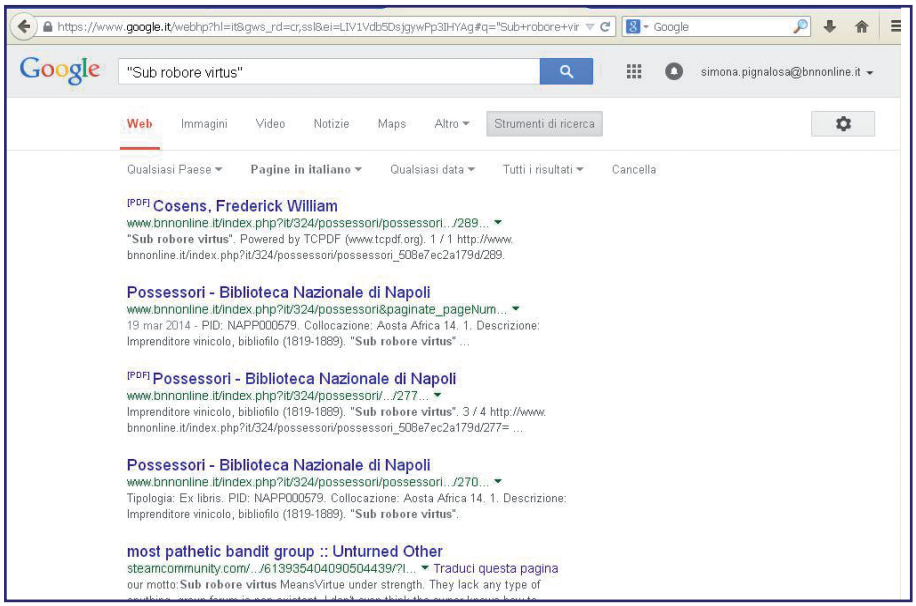


Figura 12: *Ricerca del motto in Google*

Figura 13: *La scheda nel nostro Archivio*

nienza di tali peculiarità per chi - catalogatore, bibliofilo, libraio, studioso o semplice curioso – sia impegnato nell'identificazione, non sempre agevole, di un segno di possesso. La più banale e immediata delle ricerche, la digitazione di un motto o di un acronimo in Google o in Bing, rinvia direttamente alla scheda del nostro Archivio. Lo studioso accede al nostro data base - del quale in partenza non necessariamente conosce l'esistenza - che gli offre l'opportunità insostituibile del confronto con l'immagine riprodotta, unitamente alle informazioni sull'antico

proprietario. In questo senso l'Archivio è utilizzabile anche come repertorio di riferimento, indipendentemente dalla sua natura principale di strumento per lo studio delle antiche raccolte che compongono la BNN.

Contenuti

L'Archivio possessori è una raccolta in continuo incremento, che procede in parallelo con la catalogazione dei fondi storici della BNN. Ad oggi, il data base contiene le immagini di 259 segni di possesso con le relative schede. Le 90 forme di rinvio, che consentono accessi alternativi, fanno salire il numero totale dei record a 349. Tra questi, 127 ex libris, 49 timbri di laici, enti o privati, 34 legature, 26 timbri conventuali o di ordini religiosi, 14 annotazioni manoscritte. Questi dati non rispecchiano esattamente la proporzione dei diversi segni di possesso nei fondi della BNN, ma risentono della volontà di dare la priorità ai segni riproducibili, come timbri, superlibros ed ex libris: elementi per loro natura seriali e quindi univocamente riconoscibili sia per l'utenza remota sia per i nostri catalogatori. Le annotazioni manoscritte inserite sono infatti prevalentemente quelle che, con una formu-

Figura 14: *Gesuiti*

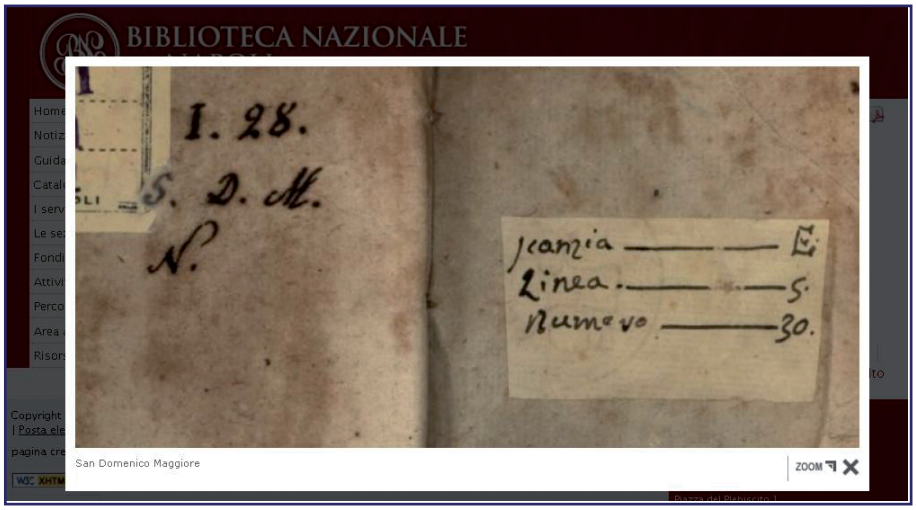


Figura 15: *San Domenico Maggiore*

Figura 16: *Legatura Farnese*

Figura 17: *Legatura alle armi dei borboni*

lazione più o meno standardizzata, suggeriscono l'appartenenza dei volumi che le contengono a nuclei di provenienza omogenea storicamente conosciuta. Esempi di queste annotazioni sono la scritta usata nella biblioteca della Casa professa dei

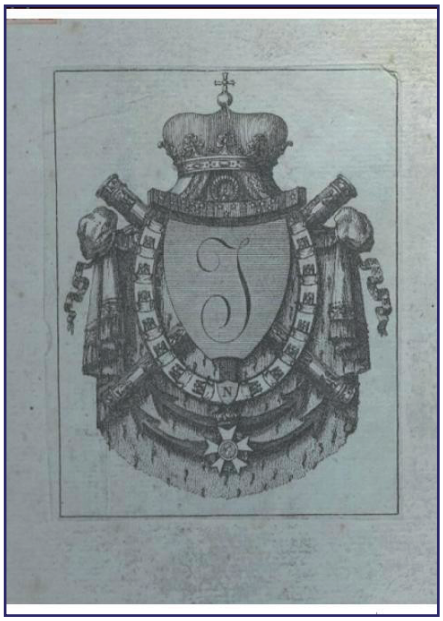


Figura 18: *Ex libris di Gioacchino Murat*

Figura 19: *Timbro della Biblioteca Brancacciana*

Figura 20: *Ex libris di Elena d'Aosta*

gesuiti napoletani (Fig. 14), o la sigla e l'antica collocazione che segnalano i libri provenienti dal convento di San Domenico Maggiore (Fig. 15).

Alcuni contrassegni sono particolarmente significativi per la storia della BNN. I libri di casa Farnese, attualmente fusi con fondi di provenienza diversa negli scaffali del fondo antico della biblioteca, sono immediatamente riconoscibili dai gigli impressi in oro sul dorso della caratteristica legatura (Fig. 16). Le dinastie che si sono succedute nei secoli XVIII e XIX sul trono di Napoli hanno decorato con il loro stemma preziose legature (Fig. 17), o hanno commissionato raffinati ex libris (Fig. 18). Nella Brancacciana, la prima biblioteca pubblica aperta a Napoli nel 1691, i libri presentano il timbro cardinalizio con lo stemma della famiglia Brancaccio (Fig. 19). L'istrice e la corona ornano uno degli ex libris usati dalla duchessa Elena d'Aosta, che donò la sua biblioteca alla Nazionale nel 1947 (Fig. 20).

Conclusioni e prospettive

Il numero di accessi all'Archivio, che si attesta intorno ai cinquecento mensili, testimonia del successo del nostro lavoro. Lavoro continuamente in progress, che si propone di completare l'inserimento dei contrassegni dagli antichi possessori schedati in SBN e non ancora testimoniati dalle relative riproduzioni. In parallelo, si incrementa costantemente il data base con l'inserzione delle immagini digitali di nuovi segni che emergono dalla catalogazione di libri antichi e non si trascura di perfezionare costantemente le schede già inserite, sia migliorando la qualità delle immagini, sia correggendo e completando le notizie, anche grazie all'apporto di studiosi che generosamente ci segnalano il risultato delle loro ricerche.

Infine, è allo studio la possibilità di creare dei link che rinviino dai record dell'Archivio alle schede dell'OPAC[6] della BNN e di dare un importante canale d'accesso al nostro data base attraverso la ricerca per provenienza nel Thesaurus del CERL(Consortium of European Research Libraries)[7].

[6] http://opac.bnnonline.it/SebinaOpacBNN/Opac.
[7] http://thesaurus.cerl.org/cgi-bin/search.pl?start=true.

L'ultima consultazione dei siti Web è avvenuta nel mese di dicembre 2015.

SISTEMA MODUS

Mostre digitali: due risultati del progetto AthenaPlus

Giuliana De Francesco - Maria Teresa Natale
Ministero per i beni e le attività culturali - ICCU

I risutati del Gruppo di studio internazionale sulle mostre digitali

Le mostre digitali sono strumenti molto efficaci a disposizione delle istituzioni culturali per presentare le proprie collezioni, raggiungendo e coinvolgendo nuovi pubblici. Se ben concepite e ben realizzate, le mostre digitali consentono di presentare diversi oggetti di interesse culturale in un comune contesto, anche quando ciò sarebbe impossibile nella realtà fisica, sviluppando intorno ad essi narrazioni digitali mirate al segmento di pubblico che si vuole raggiungere. Istituzioni culturali di diverso tipo possono collaborare alla realizzazione di una mostra digitale, in base a un obiettivo o un *concept* condiviso. Tuttavia, all'interno delle istituzioni culturali, vi è ancora scarsa competenza nella progettazione e realizzazione di mostre digitali. La ricerca nel campo è ancora limitata e non sistematica e non sono disponibili molte buone pratiche. Fa eccezione la pubblicazione da parte del Ministero per i beni e le attività culturali delle linee guida per la realizzazione di mostre virtuali online, edite nel 2011 e ristampate con aggiornamenti nel 2014[1].
Un gruppo di studio internazionale[2], nato nel 2011, sta studiando le mostre digitali dalla prospettiva delle istituzioni culturali, al fine di comprendere come raggiungere un largo ventaglio di utenti attraverso contenuti culturali attraenti e di qualità. Il lavoro di questo gruppo si basa sull'analisi dello stato dell'arte della ricerca in questo settore e intende fornire alle istituzioni culturali linee guida, risorse e altri strumenti pratici per realizzare progetti di successo[3]. Questo lavoro è stato parzialmente finanziato attraverso progetti europei, quali Linked Heritage[4] e AthenaPlus[5].
In base alla raccolta e all'analisi di un numero significativo di mostre digitali prodotte in varie parti del mondo, il gruppo di lavoro ha convenuto che la prima condizione per il raggiungimento della qualità sta nell'adozione di un approccio cura-

[1] *Mostre virtuali online: linee guida per la realizzazione*, versione 1.1, settembre 2011, integrata a luglio 2014, <http://www.otebac.it/index.php?it/320/mostre-virtuali-online-linee-guida-per-la-realizzazione>.
[2] http://www.digitalexhibitions.org.
[3] I materiali e le riflessioni sviluppati da questo gruppo di lavoro sono documentati sul relativo sito Web <www.digitalexhibitions.org> e sul sito Web del progetto AthenaPlus: <http://www.athenaplus.eu/index.php?en/200/digital-exhibitions>.
[4] Linked heritage: coordination of standards and technologies for the Enrichment of Europeana, <http://www.linkedheritage.org>.
[5] Athenaplus: access to cultural heritage networks for Europeana <http://www.athenaplus.eu/>.

SISTEMA MODUS

toriale nella concezione e implementazione di una mostra digitale. Un altro elemento importante è l'interazione con il pubblico. Il gruppo di lavoro ha pertanto elaborato una proposta di definizione di "mostra digitale", dal punto di vista delle istituzioni culturali: *"A digital exhibition is based on a clear concept and is well curated. It assembles, interlinks and disseminates digital multimedia objects in order to deliver innovative presentations of a theme, or series of themes, allowing user interaction to a great extent"*. Questa definizione identifica l'ambito in cui si muove il Gruppo di lavoro, ma è ancora in discussione all'interno del Gruppo ed è aperta al contributo della comunità professionale[6]. Per dare una solida base scientifica al proprio lavoro, il Gruppo ha raccolto una serie di risorse, tutte documentate in un database messo a disposizione della comunità professionale e degli utenti in genere. Il database fornisce l'accesso a risorse on-line o a stampa, organizzate per categorie: letteratura professionale, strumenti, linee guida, corsi di formazione, esempi e progetti. Questo repertorio viene periodicamente aggiornato ed esperti esterni al Gruppo sono invitati a contribuire.

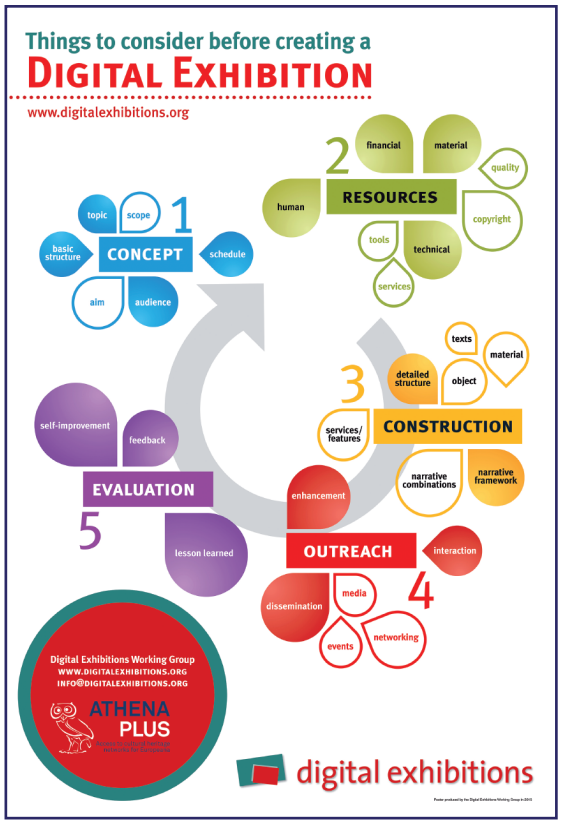


Figura 1. *Poster che sintetizza gli elementi della checklist Things to consider before creating a digital exhibition*

Uno dei recenti risultati del gruppo è costituito dall'elaborazione di una checklist, *Things to consider before creating a digital exhibition*, prodotta al fine di fornire alle istituzioni culturali un agile strumento per orientarsi nella progettazione e realizzazione di una mostra virtuale. Una prima versione di questo documento, in forma di elenco di domande con relative soluzioni, è già disponibile online[7]. Naturalmente, non tutte le domande con relative risposte si applicano necessariamente a ogni progetto, ciò nonostante affrontano temi rilevanti e forniscono sug-

[6] http://museumsdokumentation.de/joomla/digital-exhibitions/definition.
[7] http://www.athenaplus.eu/getFile.php?id=586.

gerimenti, aiutando i potenziali fruitori a orientarsi nel complesso tema delle mostre digitali. Tra i maggiori risultati di questo gruppo di lavoro internazionale, presentiamo DEMES, un set di metadati per la descrizione standardizzata delle mostre digitali. Il gruppo di lavoro era consapevole del fatto che le mostre, sia fisiche sia digitali, sono solitamente descritte dalle istituzioni culturali al fine di informare e attrarre il pubblico potenzialmente interessato. Esistono elenchi, cataloghi o inventari di mostre a livello locale, nazionale e internazionale, ma le descrizioni sono in gran parte fornite senza testo o utilizzando formati non strutturati. Il Gruppo riteneva che un formato ben strutturato per la descrizione delle mostre digitali fosse indispensabile per renderle reperibili, facilitarne l'identificazione da parte degli utenti potenziali, migliorare attraverso la standardizzazione la prassi descrittiva in uso, agevolare la condivisione dei dati e lo scambio di informazioni e includere le descrizioni delle mostre digitali nei portali locali, regionali, nazionali, europei, contestualmente a quelle degli altri contenuti culturali. Per questo motivo, il gruppo di lavoro ha avviato l'identificazione e lo sviluppo di un set di metadati per mostre digitali, a partire da standard già disponibili per la descrizione delle collezioni, considerato che la mostra digitale è un tipo specifico di raggruppamento di oggetti. Puntando alla massima diffusione possibile dello standard tra le istituzioni culturali, si è mantenuta una struttura piuttosto semplice. Gli elementi selezionati sono per la maggior parte già ampiamente utilizzati nella pratica descrittiva delle organizzazioni culturali. Il set di metadati comprende attualmente sette sezioni (identificazione, descrizione, oggetto, responsabilità, accesso, diritti, relazioni) e 30 elementi descritti fornendo ove possibile indicazioni relative al loro uso nella pratica descrittiva. Il set di metadati è stato testato dai membri del gruppo di lavoro attraverso la descrizione di mostre digitali reali. La prima bozza è stata recentemente pubblicata nell'ambito del progetto AthenaPlus[8].

MOVIO e il pilot "MUSEM"

MOVIO, un tool per costruire mostre digitali, progettato dall'Istituto Centrale per il Catalogo Unico delle Biblioteche Italiane e ulteriormente sviluppato nell'ambito del progetto europeo AthenaPlus, consente alle istituzioni culturali di valorizzare le proprie raccolte attraverso narrazioni digitali effettuate per mezzo dei diversi strumenti messi a disposizione del curatore: dallo storyteller alle gallerie di immagini, dalle timeline alle mappe interattive, agli hotspot. Il software è rilasciato con licenza MIT, la quale consente un facile riutilizzo del software da parte degli sviluppatori interessati[9].

[8] *Metadata for the description of digital exhibitions: the DEMES Element Set*, Version 0.9 (August 2015), edited by the AthenaPlus Digital Exhibitions Working Group; texts by Giuliana De Francesco (MIBACT), Arlene Peukert (SPK), Stefan Rohde-Enslin (SPK), Werner Schweibenz (BSZ), <http://www.athenaplus.eu/index.php?en/206/demes>.

[9] http://www.movio.beniculturali.it.

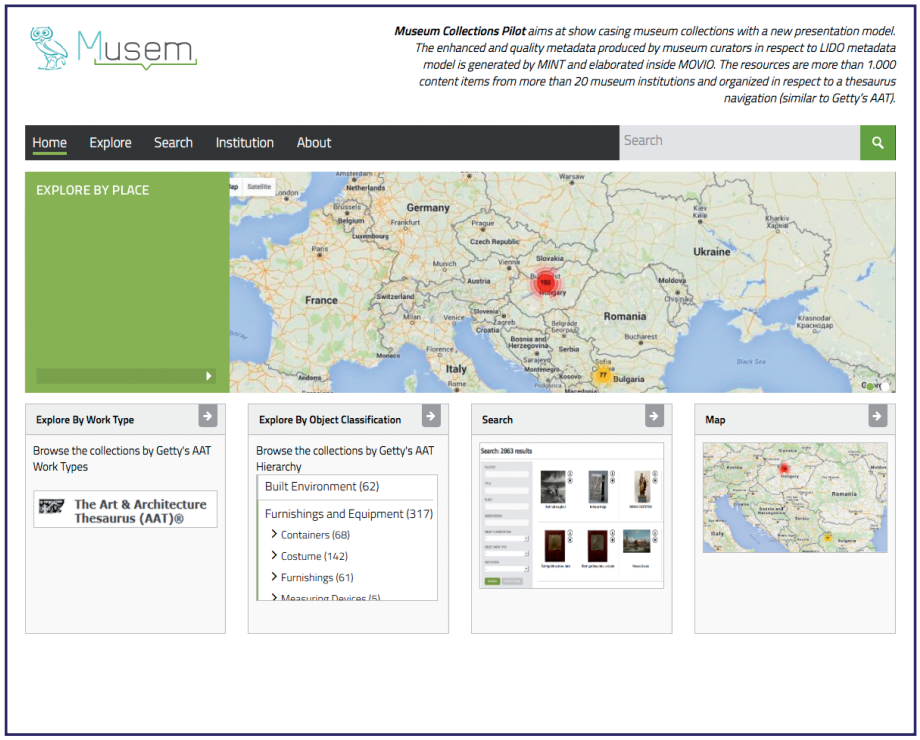


Figura 2. *Homepage del pilot MUSEM: 54.247.69.120/build/movio_museum/*

Nell'ambito di AthenaPlus è stato realizzato un esperimento di creazione di una specifica mostra digitale, sfruttando la tecnologia di MOVIO. Il pilot "MUSEM", nato dalla collaborazione tra la Philipps Universität Marburg (Germania), la National Technical University of Athens (Grecia) e il Gruppo Meta (Italia) aveva lo scopo di testare la visualizzazione di una selezione di opere museali attraverso un template appositamente studiato. Secondo la Europeana Task Force on Metadata Quality, Report and Recommendations, metadati di qualità possono essere ottenuti solo attraverso una serie di processi affidabili e tali metadati devono essere ricercabili, leggibili, standardizzati, comprensibili per l'utenza, chiari rispetto al riuso e visualizzabili[10].

Nel pilot MUSEM i metadati di qualità creati dai curatori in formato LIDO, sono stati generati da MINT[11] e rielaborati all'interno di MOVIO. Più di mille risorse digitali sono state messe a disposizione da 20 istituzioni museali e rese ricercabili at-

[10] *Report and Recommendations from the Europeana Task Force on Metadata Quality,* December 2013-May 2015, disponibile all'indirizzo: <http://pro.europeana.eu/files/Europeana_Professional/Publications/MetadataQualityReport.pdf>.

[11] MINT è una piattaforma Web progettata e sviluppata per consentire l'aggregazione di metadati di istituzioni culturali che intendono rendere le proprie risorse digitali interoperabili con Europeana, <http://mint.image.ece.ntua.gr/redmine/projects/mint/wiki/Wiki>.

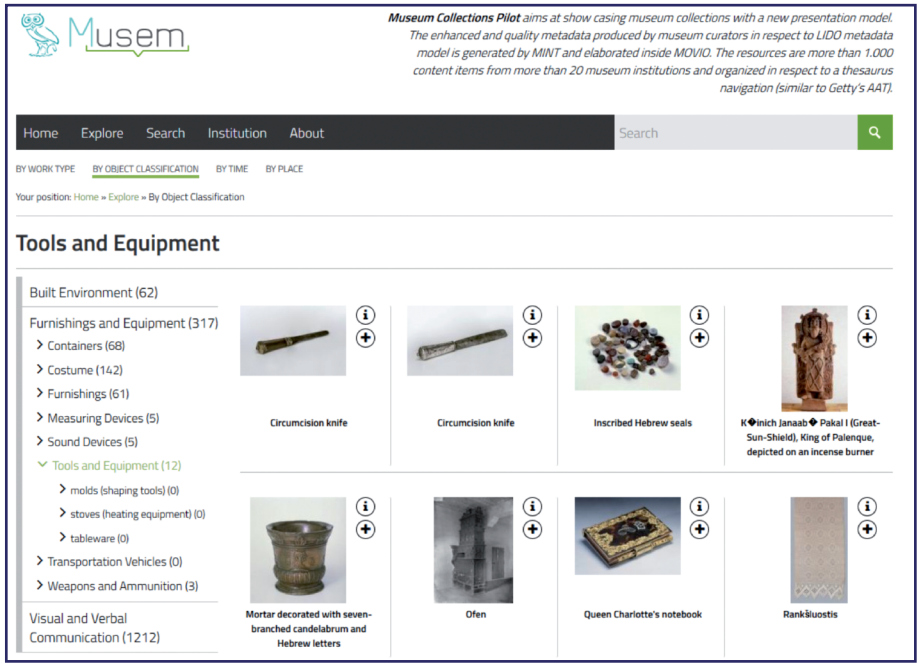


Figura 3. *Visualizzazione delle risorse digitali selezionate in base alla Classificazione per tipologia di oggetti (Strumenti e utensili)*

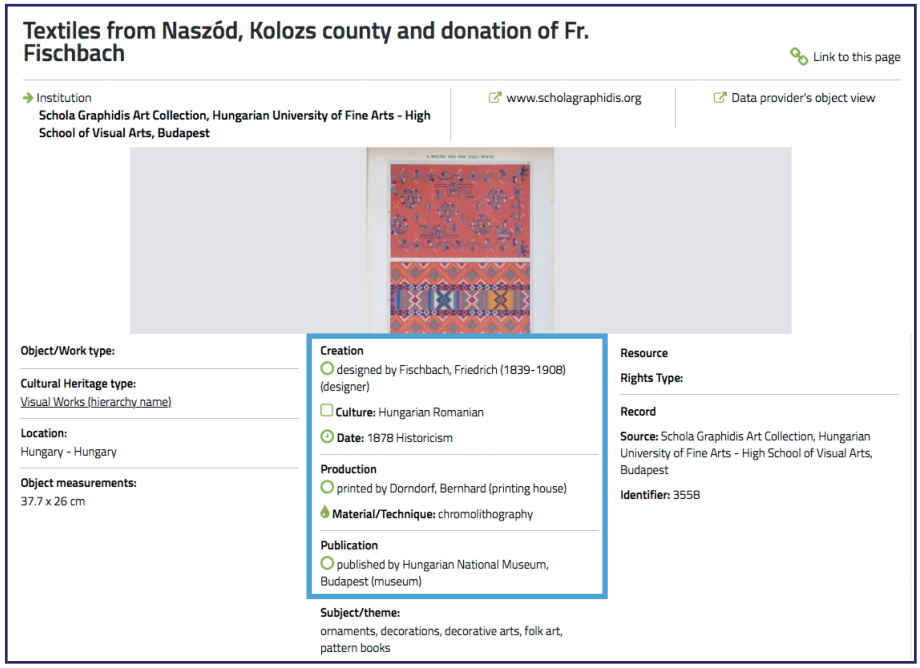


Figura 4. *Visualizzazione degli eventi relativi alla storia di un singolo oggetto*

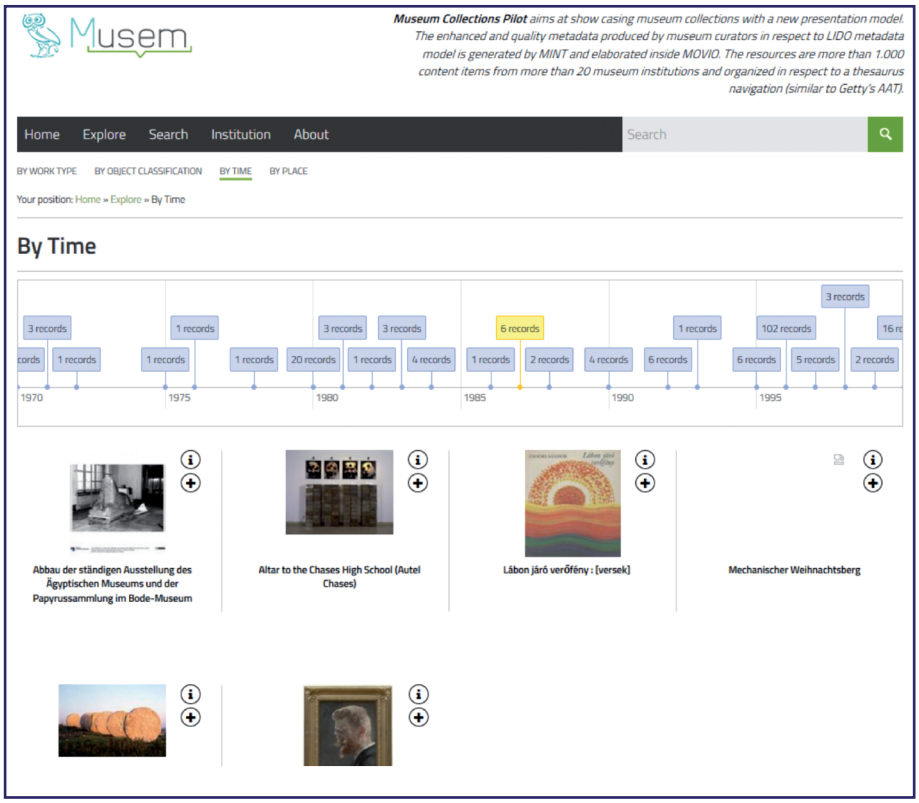


Figura 5. *Visualizzazione degli oggetti attraverso l'uso della timeline*

traverso una navigazione basata su thesauri (come l'AAT del Getty). Infatti, l'utente può cercare gli oggetti digitali per classificazione, data, luogo e istituzione.
MUSEM quindi è un esempio di mostra digitale complessa, che riutilizza gli oggetti digitali di alcuni partner del progetto e che ha consentito un'ulteriore sperimentazione del software MOVIO. Il pilot è stato il prodotto finale di una serie di requisiti formulati da una task forse dedicata che aveva l'obiettivo di contribuire alla discussione in Europeana sul miglioramento della qualità dei metadati. MUSEM non intende essere il prototipo di un nuovo portale dei musei, ma un'applicazione di MOVIO che ha l'obiettivo di riutilizzare e migliorare la ricercabilità e visualizzazione di un numero limitato di oggetti museali, e allo stesso tempo rispondere alla richiesta di qualità e miglioramento dei metodi di ricerca degli oggetti digitali in portali complessi. Una cura particolare è stata posta nella visualizzazione dei metadati, cercando di rispettare le già citate raccomandazioni di Europeana, come si può evincere dalle schermate in figura.

L'ultima consultazione dei siti Web è avvenuta nel mese di dicembre 2015.

Il Polo digitale degli istituti culturali di Napoli

Luigi Cerullo
ICCU

1. Introduzione

L'idea di un polo di istituti culturali nasce nel luglio 2013, quando l'Istituto Italiano per gli Studi Storici[1], il Pio Monte della Misericordia[2], la Società Napoletana di Storia Patria[3] e la Fondazione Biblioteca Benedetto Croce[4] decidono di cooperare - nell'ambito delle misure adottate dalla Regione Campania e destinate alla digitalizzazione di archivi e biblioteche - unendo i fondi europei destinati all'acquisto delle attrezzature tecnologiche previste dal bando "Digitalizzazione e messa in rete di archivi e biblioteche pubblici e privati". Al nucleo originario si affianca nel dicembre dello stesso anno la Cappella del Tesoro di San Gennaro[5].

1.1 I vantaggi della cooperazione

In primo luogo la condivisione delle risorse economiche, dovuta al contributo che ciascun istituito ha deciso di mettere a fattor comune, ha determinato la disponibilità di budget che ha consentito di progettare il sistema di gestione del patrimonio. La creazione del Polo ha reso possibile poi – sia grazie al prestigio delle istituzioni coinvolte sia grazie alla massa critica delle risorse potenziali – richiedere la collaborazione dell'ICCU (Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche) nella fase di sviluppo di alcune delle componenti più innovative del software gestionale: in particolare la componente semantica che consente di generare dataset RDF a partire da fonti dati SBN (vedi paragrafi *6 e 7*). Tale collaborazione è inoltre finalizzata al rilascio di una piattaforma opensource, gratuitamente disponibile, che consentirà alle istituzioni della cultura di gestire l'intero ciclo di produzione e gestione del digitale finalizzato all'erogazione di servizi destinati alla comunità dei propri utenti.

[1] http://www.iiss.it/.
[2] http://www.piomontedellamisericordia.it/home/listituzione/.
[3] http://www.storiapatrianapoli.it/.
[4] http://www.fondazionebenedettocroce.it/.
[5] http://www.cappellasangennaro.it/.

1.2 Il progetto di digitalizzazione

Il progetto è articolato in due ambiti che procedono paralleli: lo sviluppo della piattaforma di gestione (sulla quale è per lo più incentrato il presente lavoro) e l'attività di digitalizzazione del patrimonio, costituito prevalentemente da documenti d'archivio, da testi a stampa antichi e da materiale grafico.

Figura 1: *un esempio del patrimonio documentale di pregio posseduto dagli istituti culturali, composto da manoscritti, disegni, materiale d'archivio, testi antichi e opere d'arte*

I file prodotti dalla scansione ottica - prevalentemente attraverso l'uso di scanner planetari a sensore d'area - sono per lo più finalizzati ad offrire "copie per immagine" del bene; solo in taluni casi, tale processo è il primo passo verso la creazione di una vera edizione digitale dell'opera, attraverso l'estrazione automatica di testo ed entità significative e la conseguente marcatura basata sull'uso di grammatiche standard.

1.2.1 Il patrimonio oggetto di intervento in numeri

Circa 620.000 oggetti digitali nel formato master (considerando invece tutte le istanze prodotte si conteranno alla fine circa 1.800.000 file), così distribuiti in base al tipo di materiale:

- 330.000 pagine d'archivio di documentazione moderna e contemporanea
- 220.000 pagine di testi a stampa antichi
- 60.000 file relativi a manoscritti e pergamene
- 10.000 file di materiale grafico (fotografie, disegni e stampe).

1.3 Gli standard di digitalizzazione

I file vengono generati in formato TIF non compresso con campionamento a 300 ppi per tutti i documenti a stampa con dimensioni pari o superiori al formato A4; a 400 ppi per quelli inferiori a tale misura. Per i manoscritti, le pergamene e il materiale grafico (fotografie, stampe e disegni) il campionamento base sale a 400 ppi per arrivare fino ai 600 nel caso di documenti di dimensioni inferiori alla misura standard individuata. Quando le dimensioni o il particolare tipo di materiale lo richiedano, si prevede comunque di arrivare a valori di campionamento molto più elevati. Tutti i file sono prodotti con profondità colore di 24 bit RGB.

Tali file costituiscono la copia master per l'archiviazione e la conservazione. La conservazione degli oggetti è affidata a un servizio cloud di storage management che garantisce l'integrità dei file e consente, interfacciandosi con il sistema di gestione, di avviare servizi di delivery anche su tali formati.

I file derivati sono JPG con valori di campionamento che dipendono dall'uso: la piena risoluzione è destinata alla consultazione interna e alla vendita attraverso il sistema di e-commerce integrato nella piattaforma; i sotto-campionali (in media 150 ppi) saranno visibili on-line ed esposti tramite OAI-PMH[6] verso gli aggregatori (vedi paragrafo *3.2*).

2. Il sistema di gestione dei dati (Data Management)

L'architettura generale del sistema, da un punto di vista funzionale, è piuttosto complessa e prevede la presenza di numerosi moduli che interagiscono tra di loro al fine di erogare i servizi necessari a supportare i complessi processi di lavoro finalizzati alla descrizione, digitalizzazione, conservazione e divulgazione di beni culturali documentali e museali. Punto di forza della soluzione adottata è la disponibilità, in un ambiente integrato Web based, di tutti i servizi software dedicati in primo luogo alla descrizione del patrimonio e alla sua valorizzazione, poi, attraverso la gestione del ciclo di produzione e disseminazione in ambiente digitale.

2.1 I modelli dei dati

Il sistema di gestione dei dati prevede diversi moduli di catalogazione e descrizione dei materiali:

– Un modulo di descrizione archivistica basata sullo standard ISAD (G)[7] e la sua codifica XML EAD (Encoded Archival Description)[8] che rende le descrizioni prodotte con tali sistemi indipendenti dalla particolare architettura software implementata, consentendone agevolmente l'esportazione verso altri sistemi

[6] https://www.openarchives.org/pmh/.

[7] ISAD (G): General International Standard Archival Description, le cui specifiche sono disponibili all'indirizzo: <http://www.ica.org/10207/standards/isadg-general-international-standard-archival-description-second-edition.html>.

[8] http://www.loc.gov/ead/.

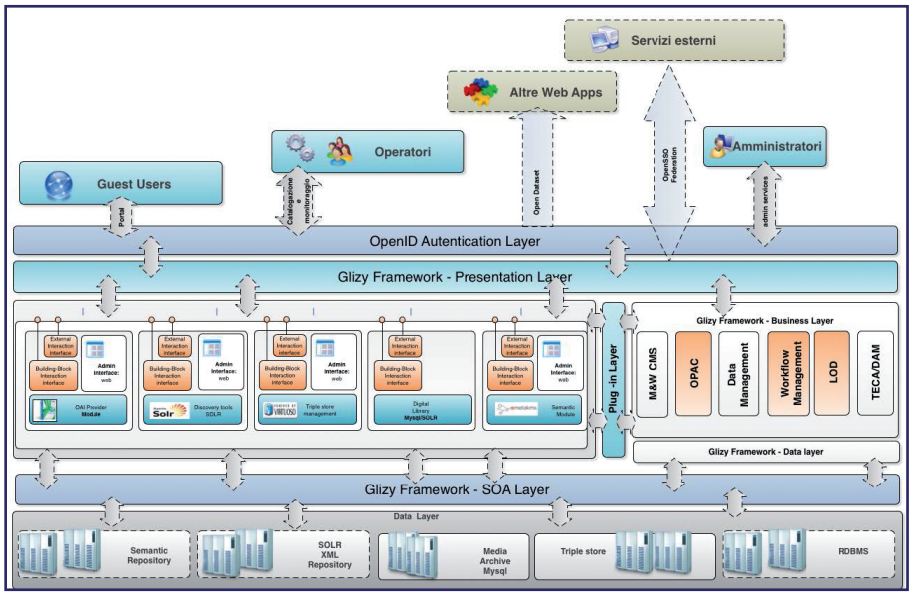


Figura 2: *diagramma funzionale dell'intera piattaforma di gestione*

che adottino tale standard per l'interscambio dei dati. In generale ogni modulo del sistema di gestione del Polo è basato su uno standard di struttura e di scambio dati che garantisce al catalogo creato la sua esportabilità su altre piattaforme. Il profilo applicativo studiato per il Polo consente la descrizione di complessi archivistici di documentazione moderna e contemporanea fino ad un livello analitico (ISAD 4, livello di descrizione dell'unità documentale).

In linea generale la catalogazione in Museo&Web FAD - sistema che rappresenta la base open-source nonché la componente funzionale core del sistema di gestione del Polo da cui si è partiti nello sviluppo, affidato al Gruppometa srl[9], partner tecnico del presente progetto - permette di utilizzare un insieme di strumenti che agevolano il compito dell'operatore: elenco dei campi gia compilati, gestione di liste e vocabolari (authority files), campi ripetibili, ecc. Tali strumenti sono forniti all'interno di interfacce grafiche di semplice utilizzo descritte di seguito.

– Un modulo che consente la catalogazione del patrimonio sulla base delle normative ICCD (Istituto centrale per il catalogo e la documentazione). In tale caso l'applicazione consente l'import-export nel formato standard definito dall'ICCD (trc 92). Le normative previste sono: F (fotografia)[10], D (disegno)[11], S (stampa)[12] e OA (oggetto d'arte)[13].

[9] http://www.gruppometa.it.
[10] http://www.iccd.beniculturali.it/index.php?it/473/standard-catalografici/Standard/10.
[11] http://www.iccd.beniculturali.it/index.php?it/473/standard-catalografici/Standard/21.
[12] http://www.iccd.beniculturali.it/index.php?it/473/standard-catalografici/Standard/33.
[13] http://www.iccd.beniculturali.it/index.php?it/473/standard-catalografici/Standard/29.

- Un modulo di integrazione con l'ambiente SBN (Servizio bibliotecario nazionale) basato su un sistema di cooperazione applicativa che consente al software del Polo di importare i dati bibliografici dal Polo SBN che fa capo alla Biblioteca nazionale di Napoli (tale Polo costituisce per i partner del progetto l'ambiente di lavoro SBN, rappresentato dal software gestionale Web-based SBNWeb[14], sviluppato dall'ICCU e usato da oltre 20 poli SBN). Tale scelta è dettata dall'esigenza di non investire risorse nella duplicazione di servizi che sono già disponibili sul territorio e ampiamente testati, oltreché dall'esigenza di rendere massima la visibilità del patrimonio librario e grafico grazie all'immediata disponibilità dei dati bibliografici resa possibile dal sistema informativo SBN.
- Un modulo che consente la descrizione dei manoscritti basata sullo standard TEI-MS (Text Encoding Initiative Manuscript Description)[15] adottato dalla piattaforma Manus-online dell'ICCU, che permette di importare le descrizioni in tale sistema informativo che rappresenta di fatto il censimento dei manoscritti posseduti dalle bibloteche italiane.

2.1.1 Schema del modulo di gestione dati con tutti i canali gestionali esplicitati suddivisi nei principali domini descrittivi

1. Patrimoniale:
   a. OA (opera d'arte)
   b. F (fotografia)
   c. D (disegno)
   d. S (stampe)
2. Archivistico:
   a. EAD (ISAD compliant): standard internazionale di codifica della descrizione archivistica
   b. EAC-CPF[16] (ISAAR CPF[17] compliant): standard internazionale di codifica per la descrizione dei soggetti produttori
3. Bibliografico:
   a. SBN Marc (in cooperazione applicativa)
   b. MODS[18] (Metadata Object Description Schema)
   c. TEI-MS (Editor TEI Manus compliant)

[14] http://www.iccu.sbn.it/opencms/opencms/it/main/sbn/SbnWeb/.
[15] http://www.tei-c.org/release/doc/tei-p5-doc/it/html/MS.html.
[16] Encoded Archival of Context, standard di struttura basato su metalinguaggio xml conforme allo standard di contenuto ISAAR: <http://eac.staatsbibliothek-berlin.de/>.
[17] ISAAR (CPF) Standard internazionale per i record d'autorità archivistici di enti, persone, famiglie
[18] http://www.loc.gov/standards/mods/.

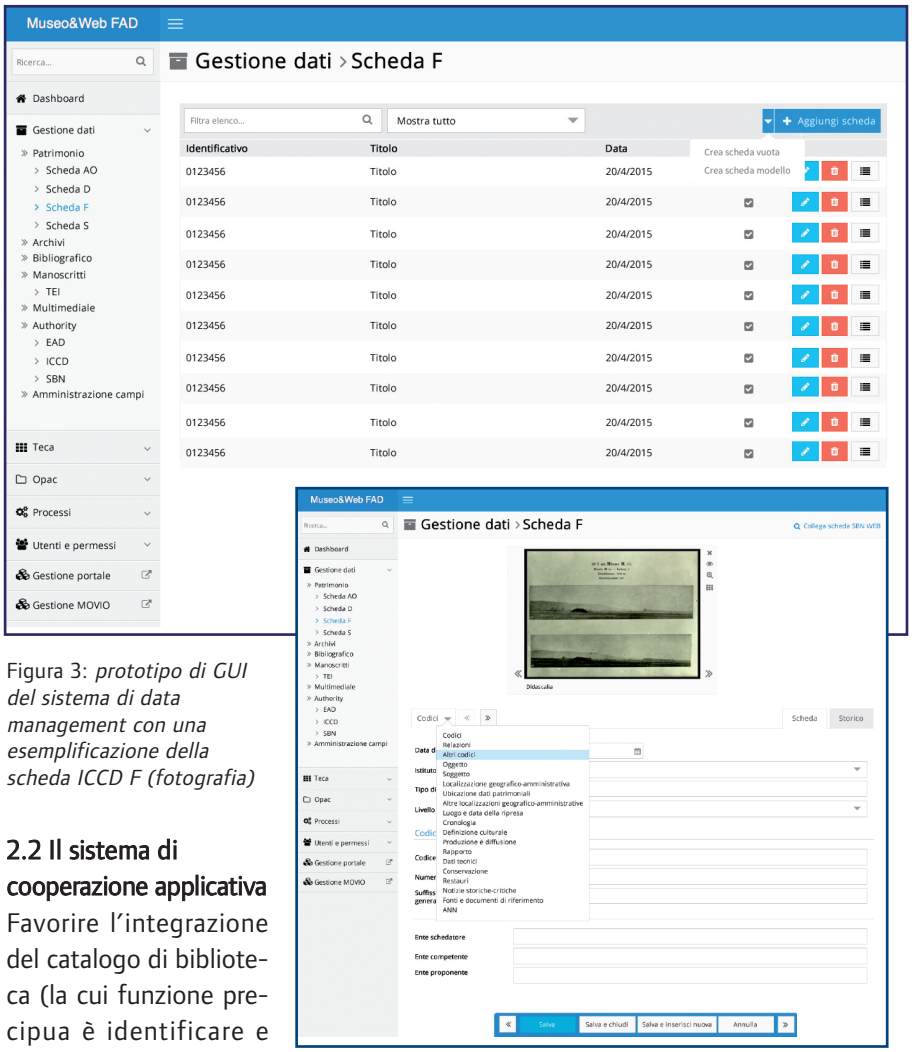


Figura 3: *prototipo di GUI del sistema di data management con una esemplificazione della scheda ICCD F (fotografia)*

2.2 Il sistema di cooperazione applicativa

Favorire l'integrazione del catalogo di biblioteca (la cui funzione precipua è identificare e localizzare una risorsa documentale) con i cataloghi che descrivono il bene culturale sulla base di modelli specialistici è uno degli obiettivi di tale progetto: in altre parole favorire l'integrazione tra gli ambienti descrittivi, in particolar modo quando questi si sovrappongono parzialmente perché restituiscono, a livelli di approfondimento diversi e per utenze differenziate, le stesse entità.

È il caso dei disegni, delle stampe e delle fotografie per le quali il sistema del Polo digitale prevede l'integrazione tra l'ambiente SBN, finalizzato alla descrizione immediatamente disponibile e fruibile attraverso il sistema informativo nazionale rappresentato dall'OPAC SBN, e i cataloghi basati sulle normative definite, per tali beni, dall'ICCD.

Per il Polo digitale, SBN infatti è l'ambiente cui è riservata la fase di record-ma-

king bibliografico (che include anche materiale grafico come stampe, disegni, fotografie): il materiale librario viene gestito integralmente in ambiente SBN e importato – attraverso uno scarico dati in formato Unimarc - allo scopo di alimentare il sistema informativo del Polo digitale. Il materiale grafico (disegni, stampe e fotografie) viene, in tale scenario, preliminarmente catalogato in SBN, per poi essere dinamicamente importato, sulla base del protocollo applicativo su cui si basa il funzionamento di SBN (SBNMarc[19]), nel sistema di gestione del Polo digitale che prevede, per tali materiali, i tracciati definiti dall'ICCD (scheda F per la fotografia, scheda D per il disegno, scheda S per le stampe). Tale derivazione presuppone lo studio di mappature tra i modelli dati adottati nei due ambienti, che è tra gli scopi di una collaborazione scientifica avviata tra l'ICCU e il Polo digitale.

2.2.1 Descrizione funzionale

Gli istituti che partecipano al progetto catalogano, inventariano e collocano il materiale grafico, come detto, con le procedure di SBNWeb, usando le normali utenze disponibili sull'ILS[20] di Polo. I record così creati sono poi estratti da Indice o da Polo in cooperazione applicativa tramite protocollo SBNMarc - attraverso una interfaccia di ricerca dedicata, disponibile in Museo&Web FAD (componente di data management della piattaforma di gestione del Polo digitale) – con l'invocazione di un servizio specifico ("cerca") tra quelli ammessi dal protocollo.

Le procedure cablate nel sistema d'arrivo, definite sulla base delle mappature studiate, consentono la precompilazione dinamica dei campi ICCD mappati con le strutture dati SBNMarc importate.

Lo scopo di tale processo è consentire, attraverso un unico flusso di lavoro, di alimentare sia il sistema informativo SBN che i repository ICCD locali, estendendo e arricchendo sul sistema di gestione del Polo digitale la descrizione dei suddetti materiali, attraverso i tracciati specialistici disponibili secondo il livello di descrizione scelto dai diversi istituti (inventariale, pre-catalogo e catalogo).

L'output finale - grazie ai formati di scambio standard previsti in fase di export dei dati - sarà un pacchetto generato in formato ICCD92 che potrà essere importato nelle banche dati regionali e centrali.

Nell'ambito di tale progetto è allo studio la fattibilità di una integrazione con il sistema SigecWeb[21] in modo analogo a quanto previsto con l'ambiente SBN. Infatti per tale sistema sono documentate API di tipo Web services[22], principalmente per

[19] Le specifiche del protocollo alla base del sistema di cooperazione applicativa attraverso il quale è realizzato il catalogo collettivo sono disponibili all'indirizzo: <http://www.iccu.sbn.it/opencms/opencms/it/main/sbn/evoluz_indice_sbn/pagina_143.html>. Il protocollo (nella versione 2.0) consente anche la gestione del materiale audiovisivo.

[20] Integrated Library System, rappresentato da sbnweb.

[21] http://www.iccd.beniculturali.it/index.php?it/118/sistema-informativo-generale-del-catalogo-sigec.

[22] http://www.iccd.beniculturali.it/index.php?it/394/interoperabilit.

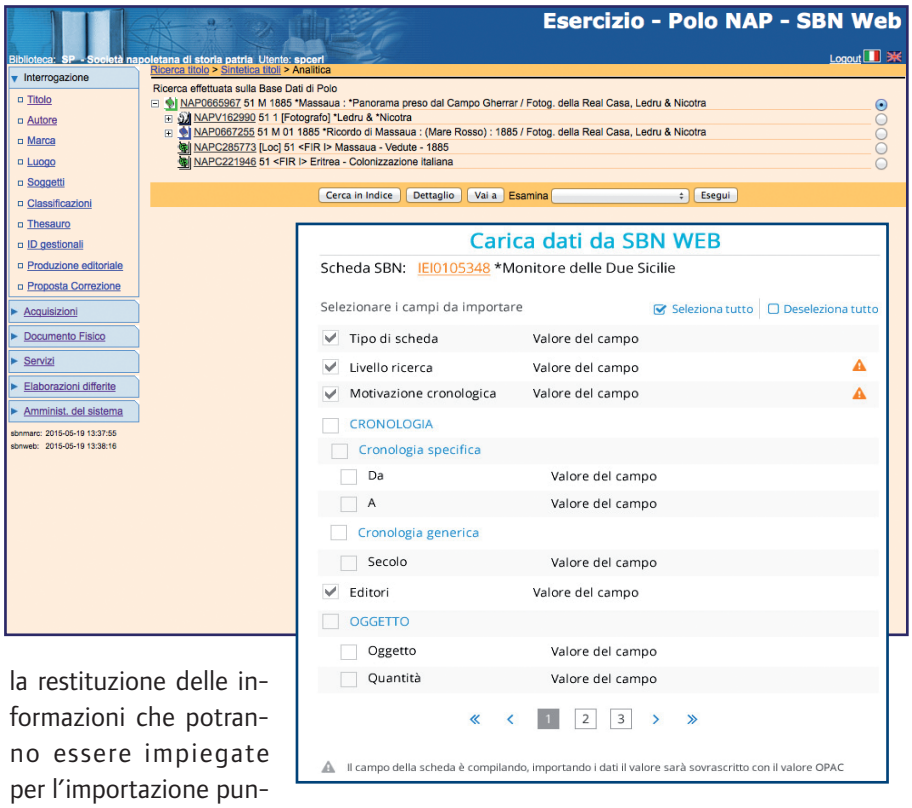


Figura 4: *nella prima immagine è rappresentato l'ambiente di recordmaking SBN costituito dal gestionale di Polo SBNWeb. Il protocollo applicativo SBNMarc consente al Polo digitale di acquisire dinamicamente i dati creati in SBN popolando, sulla base delle mappature previste, i campi della normativa ICCD corrispondente*

la restituzione delle informazioni che potranno essere impiegate per l'importazione puntuale di singoli record, mentre per l'importazione batch nel sistema locale di schede eventualmente già presenti nel SigecWeb si userà il formato di scambio ICCD92 sopra citato.

3. Digital Assets Management (DAM) e funzionalità di Teca digitale

La componente funzionale fondamentale in una piattaforma gestionale di tale natura è indubbiamente il sistema di digital library. Il progetto prevede lo sviluppo di una teca digitale che consentirà al personale di gestire i file digitali prodotti (che siano originati da campagne di scansione – come in questa prima fase – o che siano prodotti internamente sulla base di richieste di servizi – ad esempio richieste di digitalizzazione avanzate dagli utenti attraverso i canali di richiesta disponibili sul sito-portale), attraverso la creazione dei metadati gestionali. In ambiente digitale i metadati gestionali sono fondamentali per consentire, attraverso un sistema di interfacce informatiche, di ricostruire virtualmente l'integrità fisica e la struttura logica di un documento analogico e del testo/contenuto che esso veicola. Accanto

ai metadati tradizionalmente associati alla ricerca in un catalogo (valori che garantiscono le funzioni di ricerca, localizzazione e selezione di una risorsa informativa), risultano imprescindibili i metadati di natura strutturale che consentono la corretta rappresentazione della struttura fisica del documento analogico (l'organizzazione delle parti costituenti il supporto documentale) e della struttura logica dell'interfaccia testuale (ad esempio i capitoli di un libro). Tali metadati aggiuntivi vengono generati manualmente attraverso la disponibilità di editor sviluppati ad hoc.

Per garantire la massima flessibilità di utilizzo, nonché l'integrabilità con la componente di data management Museo&Web FAD, una facile manutenibilità e prestazioni adeguate, scalabilità, sicurezza e flessibilità d'uso, si è deciso di implementare una soluzione di Digital Asset Management (DAM) per la gestione degli oggetti digitali basato su framework Glizy scritto in PHP, motore di ricerca Apache SOLR e un front-end realizzato in AngularJS[23].

Caratteristica principale del DAM è la possibilità di specificare in modo estremamente accurato tutti i set di operazioni che possono essere effettuate su ogni singolo oggetto digitale come:

- inserimento, modifica e cancellazione degli oggetti digitali
- ricerca e consultazione degli oggetti complessi[24]
- editing dei metadati nei formati supportati
- gestione delle collezioni o delle raccolte per facilitarne il riuso
- tagging gerarchico dei contenuti al fine di facilitare le operazioni di ricerca
- conversione degli oggetti digitali nel maggior numero di formati compatibili
- gestione di oggetti gerarchici
- gestione di oggetti multiformato
- supporto al protocollo OAI-PMH

3.1 Profili di metadati generati e workflow di metadatazione

Le linee dedicate alla codifica dei metadati gestionali contemplano processi che è possibile definire come semi-automatici. Una volta manipolato attraverso le funzionalità del DAM, l'oggetto digitale - che sia rappresentato da un singolo file o da una collezione di file - è agganciato, attraverso apposite intefacce di ricerca, alla fonte dati descrittiva dell'entità intellettuale rappresentata dal record bibliografico o dalla scheda archivistica. Tale fonte, grazie a mappature cablate nel sistema, consente il popolamento automatico della sezione descrittiva dei metadati gestionali.

[23] Framework javascript opensource, supportato da Google. Angularjs è una infrastruttura per la creazione di applicazioni client side. Dalla documentazione ufficiale: «… Angular è quello che HTML avrebbe dovuto essere se fosse stato progettato per sviluppare applicazioni …». Per approfondimenti: <https://angularjs.org/>.

[24] Un oggetto complesso in tale ambiente è rappresentato da una associazione tra bytestreem (il vero e proprio oggetto digitale) e datastreem (metadati espliciti definiti in un qualsiasi schema).

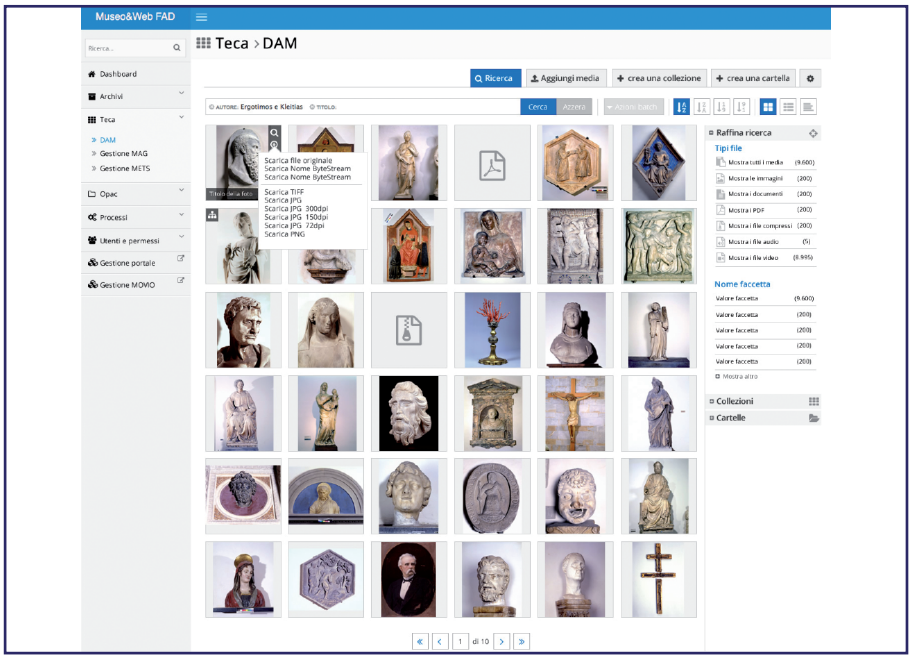


Figura 5: *generazione dei formati derivati a partire dal master conservativo, operazione effettuata su singolo oggetto attraverso le funzionalità del DAM*

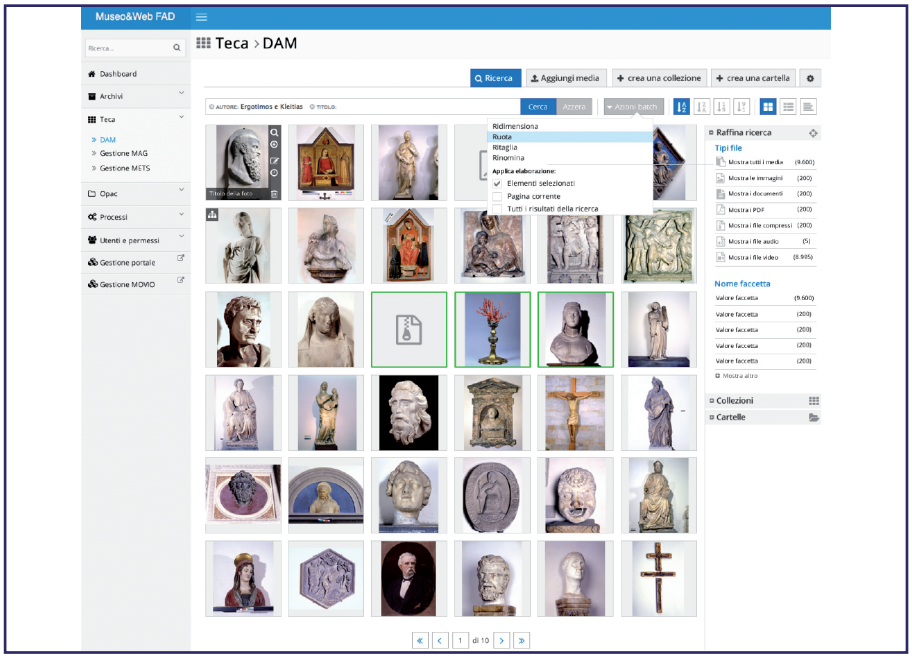


Figura 6: *avvio di procedure batch di lavorazione dei file attraverso le procedure integrate nel DAM*

Per i metadati strutturali il sistema dispone di interfacce dedicate che consentono all'operatore di stabilire sequenza e nomenclatura dei file componenti l'oggetto della metadatazione. Tali linee procedurali dispongono di un viewer che consente di visualizzare e/o riprodurre i file in modo che l'operatore sia in grado, una volta stabilitane la posizione, di definirne il contenuto attraverso i metadati di nomenclatura (elemento esplicitamente dichiarato nello schema MAG[25], che trova corrispondenza anche nel framework METS[26]).
Una volta definita la struttura fisica, l'operatore passa alla definizione della struttura logica attraverso un'interfaccia grafica che gli consente di aggiungere elementi, anche nidificati, definendone la nomenclatura e la posizione, attraverso la valorizzazione di puntatori agli oggetti digitali la cui sequenza sia stata già definita nella fase preliminare della strutturazione fisica. L'interfaccia dispone a tal fine di un viewer che consente all'operatore di visualizzare i file e definirne i puntamenti.

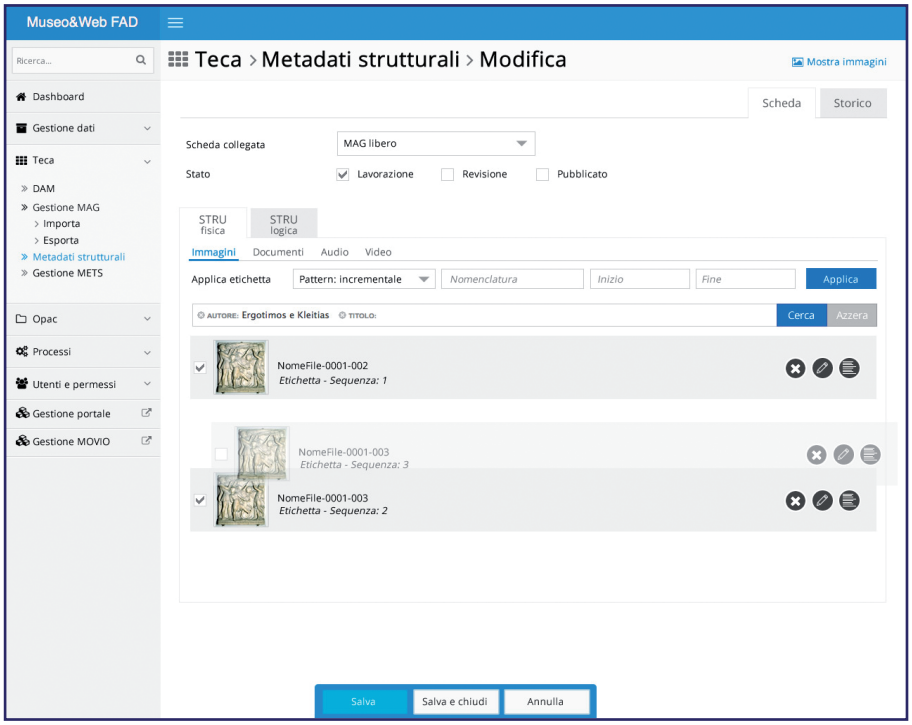


Figura 7: *interfaccia di definizione dei metadati strutturali: consente la definizione della struttura fisica di un oggetto complesso e la descrizione della sua tavola dei contenuti (struttura logica)*

[25] Metadati Amministrativi Gestionali: standard italiano di metadatazione pensato per la gestione per lo più di oggetti digitali generati da campagne di digitalizzazione. Per approfondimenti: <http://www.iccu.sbn.it/opencms/opencms/it/main/standard/metadati/pagina_267.html>.

[26] Metadata Encoding and Trasmission Standard: framework standard per la codifica dei metadati descrittivi, amministrativi e strutturali sviluppato dalla Library of Congress. Per approfondimenti: <http://www.loc.gov/standards/mets/>.

3.1.1 Schema delle componenti funzionali del sistema sotto il profilo dei servizi digitali disponibili e dei profili di metadati gestiti:

1. DAM (Digital Assets Management o sistema di Digital Library):
   a. Gestionale di risorse digitali: consente la manipolazione degli oggetti digitali e la generazione dei formati file derivati
   b. Gestionale di pacchetti (SIP AIP DIP secondo il modello funzionale e informativo OAIS[27])
2. Generatore e editor di metadati MAG e METS. Formati gestiti:
   a. MAG: consente l'esposizione degli oggetti verso l'aggregatore Internet culturale
   b. METS-SAN: consente l'esposizione dei metadati descrittivi degli allegati digitali verso il sistema informativo del SAN (Sistema archivistico nazionale)
   c. METS-MDI (MuseiD[28])
   d. METS di Google books: profilo METS generato nell'ambito del progetto Google books

3.2 Gli aggregatori con cui colloquiamo

Lo scopo primario degli schemi di metadati sopra richiamati è quello di rendere ricercabili e usabili gli oggetti digitali quando questi siano esposti fuori dall'ambiente di produzione e gestione originario, di modo che le interfacce applicative degli aggregatori esterni (nel nostro caso Internet culturale e SAN come aggregatori primari) siano in grado di rendere correttamente i pacchetti informativi. I principali schemi di scambio gestiti sono MAG, CAT SAN e METS SAN[29]: il primo consente il colloquio tramite protocollo OAI-PMH con l'aggregatore Internet culturale verso il quale verranno esposti sia i metadati che gli oggetti digitali; il secondo consente il colloquio con il Sistema archivistico nazionale limitatamente ai metadati descrittivi codificati nella grammatica EAD (Encoded Archival Description) per quanto riguarda la descrizione dei complessi documentale (record di primo livello) ed EAC (Encoded Archival of Contest) per quel che riguarda la descrizione delle entità collegate; gli oggetti digitali e i loro collegamenti ai livelli di descrizione di pertinenza sono descritti attraverso l'uso del profilo METS-SAN che fa uso per il livello descrittivo della codifica EAD mentre per le relazioni si richiama all'ontologia SAN-

[27] L'OAIS è un modello di sistema informativo aperto per l'archiviazione di contenuti informativi (Reference Model for an Open Archival Information System), elaborato dal CCSDS (Consultative Committee for Space Data System) e certificato standard ISO 14721:2003. Il documento progettuale, nella forma di raccomandazione CCSDS, è disponibile su: <http://public.ccsds.org/publications/RefModel.aspx>.

[28] Progetto del Mibac, integrato con CulturaItalia, dedicato ai musei del Paese: <http://www.culturaitalia.it/opencms/museid/index_museid.jsp?language=it>.

[29] Tracciati di scambio che consentono di veicolare rispettivamente metadati relativi alle risorse archivistiche (CAT SAN) e alle unità documentali digitali (METS SAN).

DL[30] (SAN Digital Library). Tale sistema, a differenza di Internet culturale, riceverà solo dati descrittivi e non oggetti digitali ai quali l'utente continuerà ad avere accesso attraverso la piattaforma del sistema di Polo (*sistema afferente* nella terminologia del SAN).
La piattaforma del Polo digitale consentirà - grazie ai servizi di metadatazione integrati - un colloquio nativo con i sistemi informativi dei principali aggregatori nazionali. Gli aggregatori di primo livello (Internet culturale per il materiale bibliografico e il SAN per il materiale archivistico) consentiranno di raggiungere - attraverso i canali di interoperabilità garantiti dall'infrastruttura di Cultura Italia - l'aggregatore europeo Europeana.

4. Sistema di gestione dei processi (Workflow Management)

Una componente certamente importante in ambienti di tale complessità è il sistema di modellazione e gestione dei processi di lavoro. Quello presentato sommariamente nel presente lavoro non è un software ma un "sistema di gestione": un insieme di compo-

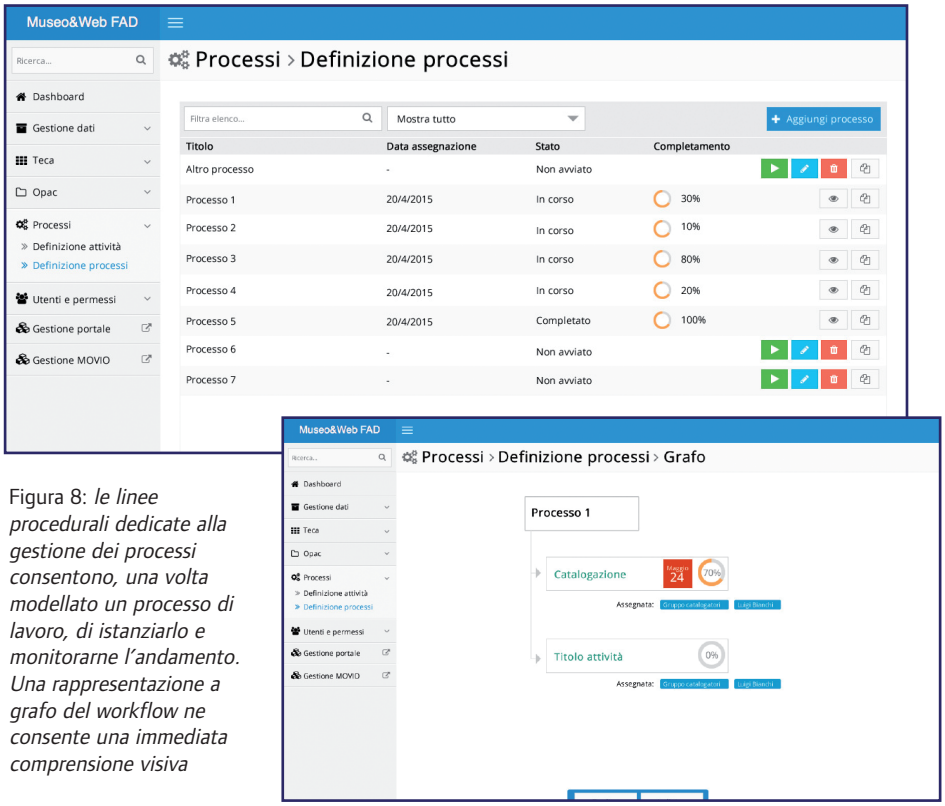


Figura 8: *le linee procedurali dedicate alla gestione dei processi consentono, una volta modellato un processo di lavoro, di istanziarlo e monitorarne l'andamento. Una rappresentazione a grafo del workflow ne consente una immediata comprensione visiva*

[30] Ontologia RDFS che consente di esplicitare, in una sezione dmdSec del METS specializzata, le relazioni che intercorrono tra i documenti digitali descritti e gli altri "oggetti archivistici".

nenti hardware, software, di risorse umane e di regole e procedure da applicare perché il flusso di lavoro finalizzato alla gestione del bene e all'erogazione dei servizi vada a buon fine e sia efficiente. Il componente di gestione dei processi di lavoro consente di modellarne i flussi ottimizzandoli e, quando necessario, ridisegnandoli (reengineering). Il servizio complesso di *document delivery*, per esempio, è gestito attraverso tale componente che consente di monitorare l'intero flusso di lavoro: dalla ricerca effettuata dagli utenti all'acquisizione delle richieste, dalla conseguente movimentazione dei materiali al processo di digitalizzazione in-house dei documenti oggetto di intervento, dalla codifica dei metadati strutturali (fisici e logici) alla disseminazione e alla vendita attraverso le componenti di accesso e restituzione della piattaforma del Polo.

5. Il portale del Polo e il viewer di risorse digitali

Per la realizzazione del portale e dei suoi servizi si useranno le funzionalità di Museo&Web CMS[31] integrate col sistema di gestione dei dati basato a sua volta su Museo&Web FAD. Grazie al CMS integrato nel sistema di data management, la fruizione dei contenuti provenienti dai cataloghi potrà essere gestita in maniera non convenzionale consentendone una più efficace rappresentazione. Il sito Web collegato al sistema di catalogazione/descrizione del patrimonio, attingendo quindi direttamente ai repositories documentali, permetterà una rappresentazione dei dati descrittivi di catalogo e degli oggetti digitali collegati sotto forma di percorsi tematici, collezioni di opere e itinerari di visita collegati a interfacce geografiche.
Il passaggio dal dato catalogato alla sua resa all'interno di percorsi di presentazione graficamente accattivanti e orientati alla didattica è certamente un punto di forza reso possibile dall'integrazione tra i due sistemi.
Un componente fondamentale in un sistema di digital library è certamente il visualizzatore di risorse digitali. Pensato come plug-in integrato nella piattaforma modulare – ma usabile anche separatamente perché basato su profili di metadati standard – si occupa del recupero e della presentazione degli oggetti digitali e dei metadati (descrittivi e strutturali) associati tramite DAM o contenuti nell'oggetto fisico lato server.
A tale componente, oltreché all'OPAC, è affidato inoltre il compito di rendere disponibili agli utenti gli strumenti necessari a effettuare le richieste di servizi sul patrimonio: quello già digitalizzato, attraverso il download integrale o selettivo dei media, a pagamento o free che siano; quello non ancora dematerializzato attraverso le richieste di digitalizzazione – effettuate direttamente da OPAC – grazie all'apposita modulistica (RIA[32]).
L'erogazione di tali servizi ha richiesto, in fase di progettazione, lo studio di un'opportuna integrazione funzionale tra le componenti applicative coinvolte nel

[31] Content Management System (CMS) pensato per facilitare la creazione e l'aggiornamento online di siti Web destinati alla valorizzazione del patrimonio e delle istituzioni culturali. Per approfondimenti: <http://www.minervaeurope.org/structure/workinggroups/userneeds/prototipo/cms.html>.

[32] Rich Internet Application: formulari dinamici online che consentono agli utenti di effettuare le richieste di servizi passando agli operatori di back-office tutti i riferimenti necessari ad avviare i

processo di acquisizione delle richieste: l'OPAC con il viewer e la modulistica RIA; questi ultimi con il server di e-commerce incluso nella fornitura.
L'interfaccia di accesso e richiesta così configurata consente di acquisire in modo dinamico i riferimenti e i parametri necessari ad avviare il processo di *document delivery*, modellato e gestito attraverso il sistema di workflow management (*vedi paragrafo 4*).

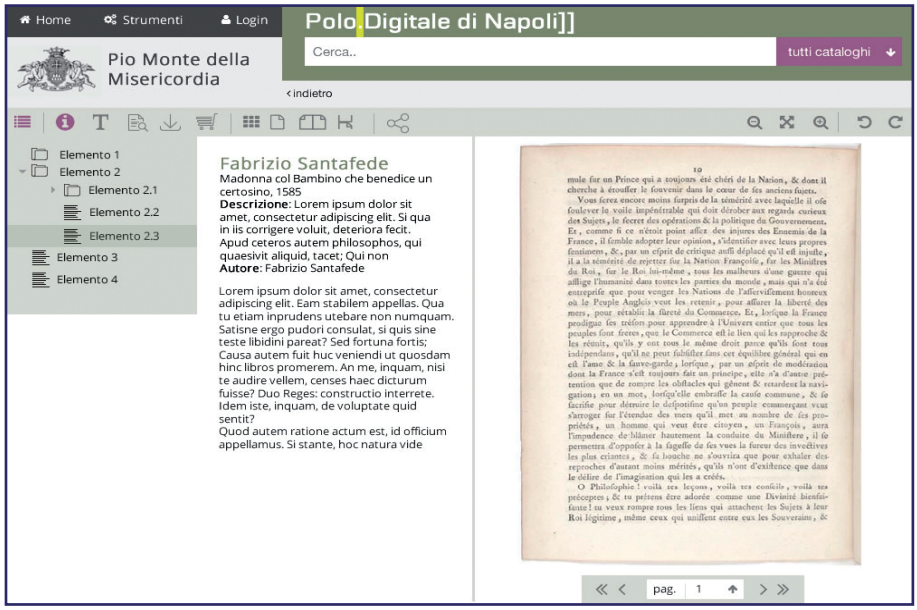


Figura 9: *il visualizzatore integrato rappresenta la componente cui è demandato non solo il compito di rendere visibili le risorse digitali, ma anche di consentire all'utenza di avviare richieste di servizi*

6. Sistema di gestione semantica e LOD (Linked Open Data)

Tra gli obiettivi più ambiziosi del progetto vi è lo sviluppo di un sistema di gestione semantica in grado di accompagnare il Polo nel complesso processo della produzione e del mantenimento dei dati aperti linkati LOD[33]. Lo scopo del sistema di gestione semantica del Polo digitale è la produzione dinamica di dataset tramite gli appositi linguaggi costituiti dal modello dei dati RDF[34] (per la strutturazione sintattica dell'informazione) e le ontologie di dominio scelte nell'ambito del progetto (come approfondito nel prosieguo a proposito dell'OPAC di Polo, il modello FRBRoo[35] per i domini bibliografico e museale e il thesauro di concetti archivistici

loro processi di lavoro, escludendo da tale flusso l'uso della posta elettronica e consentendo in tal modo al personale di risparmiare tempo, eliminando compiti ridondanti come verificare la correttezza delle informazioni riportate.

[33] http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData.

[34] http://www.w3.org/RDF/.

[35] Functional Requirements for Bibliographic Records - object oriented: <http://www.cidoc-crm.org/frbr_inro.html>.

OAD[36] per il dominio archivistico). I dataset generati – a partire dalle fonti dati interne o esterne al sistema (come nel caso dei dati SBN recuperati in cooperazione applicativa) e modellati sulla base delle ontologie di dominio – saranno collegati a dataset esterni come Dbpedia[37] (la versione opendata di Wikipedia, ad esempio per le schede biografiche di autori che potranno essere in tal modo dinamicamente richiamate), GeoNames[38] (per la georeferenziazione dei luoghi, riferibili ai luoghi di edizione di una pubblicazione, ma anche ai luoghi citati nei documenti digitalizzati) e VIAF[39] (Virtual International Authority File per le intestazioni di authority).

6.1 Arricchimento semantico e NER (Named Entity Recognition)

L'arricchimento semantico sarà utilizzato per collegare alcuni metadati contenuti nelle schede di catalogazione con risorse esterne (le ontologie dei principali dataset sopra richiamati) che potranno consentire approfondimenti o una navigazione più efficace. Questi collegamenti saranno aggiornati in tempo reale e consentiranno di avere una informazione sempre allineata alle fonti esterne. In sostanza il sistema di *semantic enrichment* analizzerà il contenuto del metadato con tecniche di NLP (Natural Language Processing), lo normalizzerà e creerà un metadato RDF che sarà esportato come tripla utilizzando una proprietà specifica come *sameAs*, il collegamento alla risorsa esterna sarà risolto in tempo reale attraverso l'URI della risorsa stessa. Tale tecnica consentirà di alimentare con un processo automatico il linking esterno con i principali dataset sopra richiamati.
Le tecniche di *semantic enrichment* saranno applicate anche al testo trascritto, o all'OCR, con lo scopo di estrarre dall'informazione non strutturata quei metadati che potranno essere usati per descrivere la risorsa in modo semiautomatico.

7. OPAC (Online Public Access Catalogue) e base di conoscenza

Pur essendo tali domini formalmente separati e gestiti sulla base di normative specifiche (SBN: per il materiale bibliografico; SAN - ISAD: per la documentazione d'archivio; normative ICCD per la descrizione del patrimonio artistico), sono diverse le relazioni e, in parte, le sovrapposizioni tra gli oggetti che costituiscono tali domini: gli archivi, residuo e traccia documentale dell'attività istituzionale di un soggetto produttore, possono testimoniare della formazione delle raccolte d'arte; le opere d'arte sono l'oggetto di studi che costituiscono intere bibliografie. Tale modello ontologico costituisce la base di conoscenza imprescindibile per costruire un punto di accesso unificato, il cui modello è rappresentato dalle ontologie di dominio FRBRoo e OAD (di cui tratteremo brevemente più avanti), impiegate nel progetto.

[36] Ontology Archival Description: <http://labs.regesta.com/progettoReload/wp-content/uploads/2013/08/oadNew.html>.
[37] http://it.dbpedia.org/.
[38] http://www.geonames.org/.
[39] https://viaf.org/.

7.1 La comunità degli utenti (interni ed esterni):

Per chi impiantare un sistema di questo tipo? Due sono le classi di utenti:

1. gli utenti esterni: gli studiosi, gli studenti, i lettori e i cultori che orbitano normalmente intorno a tale tipo di istituti culturali. Questa comunità di riferimento è caratterizzata da esigenze informative molto differenziate:
   a. le esigenze specialistiche dello studioso sono esaudite dai cataloghi di dominio: SBN per la ricerca bibliografica finalizzata all'identificazione e localizzazione dei documenti a stampa (tipica funzione del catalogo); i tracciati (di livello inventariale, pre-catalogo e catalogo) definiti dalle normative ICCD adottate per la descrizione del patrimonio, che consentono di accedere a una descrizione più dettagliata, più analitica del bene (finalizzata alla sua conservazione); e, infine, la descrizione dei complessi archivistici la cui strutturazione gerarchica, analiticità e ricchezza informativa (il modello dati è basato sullo standard internazionale ISAD) garantisce la contestualizzazione dell'informazione veicolata dalla documentazione d'archivio.
   b. le esigenze informative dell'utente generico sono esaudite dal sistema di meta-indice: identificare risorse informative di qualsiasi natura correlate in qualsiasi modo a un oggetto, un evento, una persona, una famiglia, un ente. È a questo tipo di ricerche integrate che si rivolge l'interfaccia di ricerca *cross-domain* dell'OPAC.
2. Le esigenze degli utenti interni o operatori delle istituzioni coinvolte: bibliotecari, archivisti, addetti di sala, curatori, dal lavoro dei quali dipende la gestione dei servizi all'utenza. L'introduzione della piattaforma di Polo consente l'ottimizzazione e, dove necessario, la reingegnerizzazione dei processi di lavoro in essere. Il *reengineering* dei processi è infatti uno degli obiettivi primari del progetto (*vedi paragrafo 4. Sistema di gestione dei processi*).

7.2 Un'interfaccia di ricerca innovativa (FRBR) e l'orientamento ai servizi

Nello scenario attuale il lettore riceve, in risposta a un'interrogazione, un elenco di record bibliografici che descrivono pubblicazioni sulle quali bisogna poi operare una selezione, talvolta complessa, per individuare le informazioni di interesse: gli oggetti che un sistema tradizionale restituisce sono tutti della stessa natura, sono edizioni pubblicate di date opere.

Il nuovo modello di OPAC intende rispondere in modo più puntuale alla richiesta di ciascun utente; la risposta mirata è un vantaggio che deriva dall'adozione della struttura dei dati organizzata sulla base del modello FRBR (Functional Requirements for Bibliographic Records), che distingue le diverse entità – attraverso le quali è organizzabile concettualmente il dominio della conoscenza bibliografica e museografica – e che corrispondono all'approccio della ricerca: un utente, infatti, può essere interessato ad avere informazioni su un'opera (il contenuto artistico o intellettuale identificato da una concatenazione di titolo e autore); può essere interessato a una particolare

espressione dell'opera (cioè a un determinato testo, es. una certa traduzione o una determinata edizione dell'opera); oppure può ricercare una particolare manifestazione bibliografica (l'oggetto dei cataloghi tradizionali) della quale conosce dati precisi, come il titolo, il nome dell'autore, la data di pubblicazione ecc. Questa puntualità nella risposta è possibile perché lo standard guida cui ci riferiamo prevede la produzione granulare di metadati abbinati alle entità del modello: opera, espressione, manifestazione, item; oggetto, concetto, evento, persona, famiglia, ente.

7.2.1 Schema delle componenti funzionali del sistema di OPAC

Il sistema prevede l'integrazione dei seguenti componenti software:

1. *Sistema di indicizzazione delle schede secondo i seguenti modelli dei dati:*
   a. EAD-EAC (dominio archivistico)
   b. SBNMarc (dominio bibliografico affidato per la fase di record making all'ambiente SBN)
   c. ICCD (F, D, S, OA)
   d. MODS
   e. TEI-MS (Manus compliant)
2. *Sistema di costruzione dell'indice di metaopac*
   Questo componente è l'interfaccia di amministrazione in Museo&Web FAD delle modalità di fruizione dell'OPAC per ogni tipologia di scheda e per il profilo comune. È possibile definire sia i campi di ricerca per la modalità "google like" che quelli per la ricerca avanzata.
3. *Modulo di ricerca a faccette*
   I risultati delle ricerche, sia quelle semplici che avanzate, saranno presentate corredate di faccette che consentiranno di navigare i risultati apponendo dinamicamente gli opportuni filtri definiti sulla base dei modelli gestiti.
4. *Visualizzazione sintetica e dettagliata*
   Per ogni tipologia di scheda si prevede una visualizzazione dei dati in forma sintetica che consentirà di presentare i dati più rilevanti insieme all'anteprima del documento (se presenti oggetti digitali collegati). Tutti i dati associati all'oggetto saranno visibili poi in pagine di dettaglio che ne offriranno una vista analitica organizzata per sezioni.
5. *Modulo FRBR*
   Tale componente costituisce il GUI layer (livello di presentazione dei dati) che consentirà agli utenti una navigazione del dataset RDF costruito sulla base del modello ontologico FRBRoo a partire dalla fonte dati UNIMARC di SBN. Tali interfacce consentiranno una navigazione a percorsi pre-definiti tra le entità del modello.

Le interfacce di amministrazione saranno realizzate in javascript con la tecnologia Angularjs, il motore di ricerca sarà sviluppato in SOLR e i servizi da esporre per la gestione delle interfacce di consultazione in PHP.

8. Conclusioni e prospettive

Obiettivo immediato del progetto è la creazione di un ambiente gestionale integrato che supporti le attività di valorizzazione[40] del patrimonio culturale degli istituti partner. Tali attività sono certo rese possibile dall'implementazione della piattaforma software; ma in primo luogo richiedono la formazione del personale sui complessi processi di lavoro finalizzati alla digitalizzazione e divulgazione di beni culturali documentali e museali.

Tra **ICCU** e **Polo digitale degli istituti culturali di Napoli** – come accennato in apertura del presente intervento – è attiva una collaborazione che va avanti dal febbraio 2015. Tale cooperazione, dall'intento originario limitato allo sviluppo in formato LOD dei dati bibliografici localizzati in SBN dagli istituti partner del Polo digitale, si è estesa al rilascio di una piattaforma software, gratuitamente disponibile, che consentirà alle istituzioni della cultura di sostenere l'intero ciclo di produzione e gestione del digitale, finalizzato all'erogazione di servizi destinati alla comunità dei propri utenti.

Tale obiettivo è stato formalizzato nella **Convenzione**[41] firmata nel gennaio 2016, nella quale sono stati definiti gli impegni reciproci. Da un lato la collaborazione tecnico-scientifica dell'ICCU che si esprime sia nel lavoro di estensione e raffinamento delle mappatura tra il modello dati di SBN e l'ontologia di dominio (FRBR) adottata per lo sviluppo in LOD dei dataset bibliografici, sia nel garantire le evolutive applicative dell'ambiente SBN (protocolli e servizi software) utili a favorire il colloquio in cooperazione applicativa, in particolare tra SBNWeb e Museo&Web FAD. Dall'altro l'impegno del Polo digitale che si manifesta attraverso il coinvolgimento dell'ICCU nella fase finale dello sviluppo e dell'ottimizzazione della piattaforma e nella garanzia della piena disponibilità della stessa in modo che l'ICCU possa sia riusarla a livello centrale che promuoverne la diffusione verso tutte quelle istituzioni interessate a implementare un proprio sistema di biblioteca digitale.

[40] Da intendersi qui non come monetizzazione del bene – sebbene sia disponibile anche un sistema di e-commerce – ma come accesso e conoscenza che presuppone la descrizione, lo studio e, non ultimo, la tutela del bene.

[41] http://www.iccu.sbn.it/opencms/export/sites/iccu/documenti/2016/Convenzione_ICCU_polo_digitale_istituti_cilturali_Napoli.pdf

L'ultima consultazione dei siti Web è avvenuta nel mese di dicembre 2015.

La banca dati del manifesto politico

Siriana Suprani
Fondazione Gramsci Emilia-Romagna

La banca dati *www.manifestipolitici.it* nasce da un progetto della Fondazione Gramsci Emilia-Romagna di oltre dieci anni fa. Propone alla consultazione on line una raccolta di circa 13.000 (in progressivo aumento) manifesti politici e sociali dal Novecento ad oggi. Passano sullo schermo del computer le immagini dei grandi fogli di carta che tappezzavano i muri e che, con la costituzione dei partiti di massa nel secondo dopoguerra, sono stati strumento di propaganda e di comunicazione dei partiti e dei movimenti politici.

Ho usato indifferentemente i due concetti, quello di propaganda e di comunicazione. Gli storici si sono posti invece la domanda se il loro significato sia analogo e cioè se il termine propaganda debba avere solo un'accezione negativa e alluda solo a «pratiche politiche di tipo manipolativo e prevaricatorio. Ciò ha determinato un abbandono del termine a favore di altri quali pubblicità, marketing, relazioni pubbliche o anche più semplicemente comunicazione politica»[1]. C'è del vero anche se io penso che se ci riferiamo ad un ambito politico e sociale dove vigono dinamiche democratiche e pertanto dove più voci competono e si confrontano, con libertà di decodificazione dei meccanismi, il termine propaganda sia solo storicamente datato e si possa considerare equivalente nella sostanza a quello di comunicazione politica. In ogni caso rinvio per approfondimenti al volume *Propagande contro* curato da Andrea Baravelli per i caratteri della casa editrice Carocci.

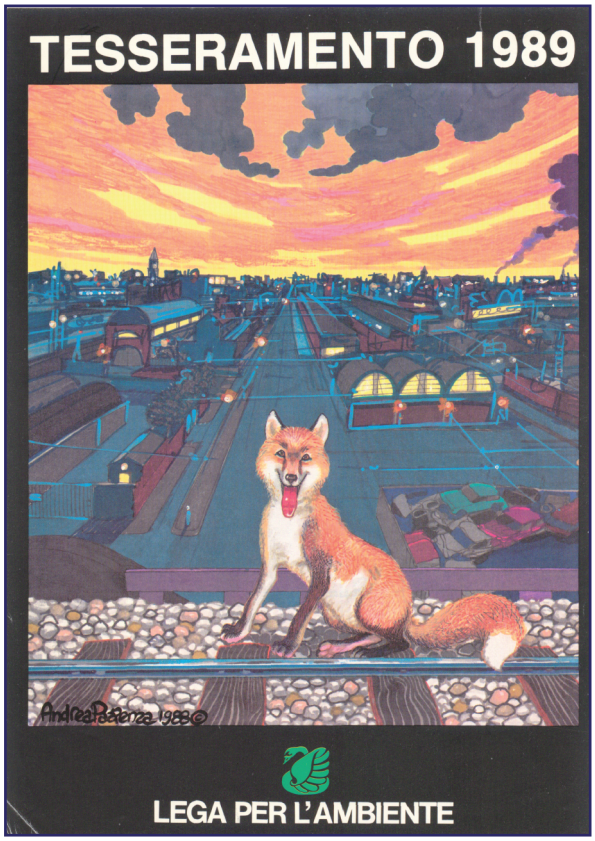
Il manifesto è stato uno dei media più utilizzati soprattutto prima dell'affermarsi della televisione. Molti documenti, come ad esempio *Il quaderno dell'attivista* periodico del PCI dei primi anni del secondo dopoguerra, testimoniano come al militante di partito si insegnasse come realizzare un manifesto, dal linguaggio all'immagine, come gli venissero indicati i luoghi più idonei per l'affissione e come scegliere i destinatari.

Il manifesto contemporaneo, soprattutto quello nel pieno della sua colorata espressione che va dal dopoguerra agli anni ottanta, è solo in parte l'erede dei bandi e dei fogli volanti, che presero a circolare fin dai tempi dell'avvento della stampa per diffondere informazioni appelli e inviti, e si affermarono in Italia sotto l'influenza della rivoluzione francese e della occupazione napoleonica[2].

[1] Andrea Bravelli (a cura di), *Propagande contro. Modelli di comunicazione politica nel XX secolo*, Roma, Carocci, 2005.

[2] *Fogli volanti toscani: catalogo delle pubblicazioni della Biblioteca di storia moderna e contemporanea di Roma 1814-1849* a cura di Sara Mori, Milano, Franco Angeli, 2008.

SISTEMA MODUS



Manifesto di Lega per l'ambiente per il tesseramento 1989; disegno di Andrea Pazienza, Roma 1988-1989.

Con essi il manifesto politico del Novecento condivide sicuramente il messaggio di breve durata, il carattere effimero dell'annuncio, legato all'evento, e la destinazione ad un largo pubblico che deve poter raccogliere il messaggio con facilità, venir raggiunto nel corso della normale attività quotidiana, quasi per caso.

Molti dei manifesti della banca dati manifesti politici.it sono dotati di immagine, di colore, di molto colore, oltre che di testo scritto. Il colore ha caratterizzato le lotte politiche del Novecento, soprattutto quelle del secondo dopoguerra, le campagne elettorali, le battaglie per i diritti civili, gli scontri e le polemiche tra i partiti. Ogni partito o movimento si distingueva anche attraverso il messaggio cromatico, grafico e letterario del manifesto, come è descritto dai saggi raccolti da Stefano Pivato e Maurizio Ridolfi nel volume *I colori della politica*[3].

Strumento per attrarre l'attenzione su un'idea, per affermare un messaggio facile che arrivi diretto dai muri della città alle persone che passano, rivolto non ai pochi ma ai tanti che poi dovranno esprimere un voto o farsi comunque un'opinione, studiato per colpire emotivamente facendo leva sul colore, sui simboli, sulla storia e sulle storie che spesso evoca, sulle immagini fotografiche e sui tratti grafici, di frequente opera di professionisti o di noti artisti. Dudovich, Steiner, Guttuso, Tedeschi, Beltrame, Jacovitti, Vitale, sono ad esempio tra le firme che si incontrano di frequente.

Mentre in *1984* George Orwell, descrive il manifesto di un regime autoritario come veicolo di passiva suggestione, tanti scrittori e letterati lo citano nelle memorie

[3] *I colori della politica: passioni, emozioni e rappresentazioni nell'età contemporanea*, a cura di Stefano Pivato e Maurizio Ridolfi., Università degli studi della Repubblica di San Marino, 2008.

SISTEMA MODUS

autobiografiche inserendolo nel racconto dei ricordi e delle emozioni, forma di narrazione ripresa da registi di teatro e di cinema che attraverso le riproduzioni dei manifesti creano il clima e l'ambientazione di un passato recente.
Il manifesto politico e sociale del Novecento è inscindibile dal contesto urbano, la sua presenza esplode nei primi decenni del secondo dopoguerra, contemporaneamente al forte processo di inurbamento di quegli anni. È di grande interesse dunque accompagnare la visione del documento-manifesto con le immagini fotografiche delle città con i muri tappezzati. Negli anni cinquanta e sessanta le macchie di colore prodotte dall'attacchinaggio spontaneo dei militanti politici di tutti i partiti allora protagonisti non erano meno invasive e senza regole di quanto non lo siano oggi i graffiti o i dipinti della street art .
Giovanni De Luna nel saggio *La politica sui muri. I manifesti del '68* scrive: «Già a partire dai primi giorni di maggio i muri del Quartiere latino cominciarono a parlare il linguaggio della politica. Erano parole di carta , affidate ai manifesti, agli slogan, alle scritte, ma erano parole che si facevano ascoltare, in grado di infrangere il silenzio della stampa e degli altri mezzi di comunicazione»[4].
La banca dati del manifesto politico non contempla solo la collezione della Fondazione Gramsci Emilia-Romagna, presente con circa 5.500 manifesti dei circa 12.000 posseduti, ma propone manifesti di tutti i partiti politici italiani e di molte associazioni, provenienti da diverse collezioni, come ad esempio quella dell'Istituto storico di Modena con i manifesti di Gino Boccasile del ventennio fascista o della Biblioteca dell'Archiginnasio di Bologna con i manifesti delle elezioni del 1948. In questi dieci anni sono state inserite collezioni di istituti di conservazione pubblici e privati, di singoli collezionisti, di raccolte amatoriali presenti su tutto il territorio nazionale e il numero è in continua espansione.
La banca dati dei manifesti è costantemente aggiornata anche sull'attualità. Il manifesto politico, dopo un periodo di relativo appannamento, è tornato a parlare, anche in modo forse troppo invadente, seppur con modalità comunicative diverse dal passato. Infatti sono soprattutto i volti dei protagonisti delle competizioni elettorali quelli che ci seguono periodicamente dai bordi delle strade. Rinviamo al saggio di Fabrice D'Almeida dal significativo titolo *Vendere meglio la propria immagine: mimica facciale e seduzione da parte dei parlamentari francesi e italiani*[5].
La collezione più importante ospitata in banca dati è quella dei manifesti della Biblioteca di Storia Moderna e Contemporanea di Roma. La raccolta proviene da una collezione privata acquisita di recente dalla Biblioteca, un insieme di oltre

[4] Giovanni De Luna, *La politica sui muri, i manifesti del '68*, in: *L'arte per la strada. I manifesti del maggio francese,* Giulio Bolaffi editore, Torino, 2008, p.47-58.
[5] Fabrice D'Almeida, *Vendere meglio la propria immagine: mimica facciale e seduzione da parte dei parlamentari francesi e italiani* in Andrea Bravelli (a cura di), *Propagande contro. Modelli di comunicazione politica nel XX secolo*, Roma, Carocci, 2005.

novemila testimonianze iconografiche di propaganda politica e sociale di associazioni, partiti politici, sindacati e movimenti; di questo insieme, più di 3.000 esemplari sono oggi visibili in *Manifestipolitici.it* attraverso la descrizione catalografica e l'immagine. Sono consultabili moltissimi documenti prodotti dalle associazioni sindacali, in particolare Cgil, Cisl e Uil e poi manifesti prodotti da tutti i partiti politici italiani, dalle associazioni culturali e ambientaliste, dai centri sociali, dai movimenti giovanili, dai movimenti delle donne, ad ogni livello territoriale, coprendo un arco cronologico che va dal 1943 agli anni 2000 e descrivendo un ricco itinerario che tocca i principali avvenimenti politici del nostro paese, dal dopoguerra ad oggi, le campagne elettorali, le campagne referendarie, le lotte e le agitazioni sociali. Questa ricca raccolta di manifesti, locandine e volantini nacque anche per conservare e valorizzare l'attività, nell'ambito della comunicazione politica, di famosi artisti e grafici italiani che anche attraverso il manifesto politico hanno voluto testimoniare il proprio impegno politico ed intellettuale: tra questi sono particolarmente rappresentati Nani Tedeschi, Ennio Calabria, Renato Guttuso, Gianni Sassi e Ettore Vitale.

Un'altra collezione che merita di essere citata è quella dei manifesti raccolti da Danilo Montaldi (1929-1975), scrittore, intellettuale, politico e consulente di diverse case editrici italiane, donati alla Fondazione Gramsci Emilia-Romagna nel 2011. Si tratta di una raccolta di 34 manifesti, relativi ai movimenti studenteschi e operai degli anni sessanta, molto interessanti anche da un punto di vista grafico. I temi presenti nei manifesti sono la lotta contro l'imperialismo americano, la guerra del Vietnam, le manifestazioni studentesche a Parigi e a Berlino negli anni della contestazione, la rivoluzione cubana e Fidel Castro insieme a documenti della Lega tedesca degli studenti socialisti, del Partito operaio di unificazione marxista spagnolo, del Movimento del 22 marzo (1968) di Daniel Cohn Bendit, di Avanguardia Operaia.

Il manifesto contemporaneo, forse più di altri documenti, ha corso il rischio della progressiva scomparsa nel parziale disinteresse, almeno fino a qualche decennio fa, da parte degli enti conservatori e della comunità degli studiosi. Le ragioni possono essere diverse. Il manifesto è un documento effimero, lo è per la sua stessa funzione di breve durata nel tempo, ma lo è anche per la fragilità e deperibilità del supporto cartaceo. Il manifesto, inoltre, non è unico e irripetibile come lo è un'opera d'arte pur esprimendo non di rado espressioni grafiche di valore avvalendosi del lavoro di menti creative. Nell'epoca della riproducibilità tecnica, per usare le parole di Benjamin, è il documento che, ancora più delle immagini fotografiche, si discosta dall'idea di unicità e appartiene invece alla riproduzione quantitativa per essere fruito in più luoghi, destinato alla visione di tanti. Non gode dell'aura dell'unicità, non solo dell'opera d'arte, ma neppure del documento d'archivio. Inoltre la sua conservazione e il corretto trattamento, a confronto con altri materiali documentari, comportano oneri maggiori.

Decidere di trasformare un accumulo di carte in una collezione di documenti a disposizione della ricerca è un atto che chi fa il mestiere dell'archivista o del bibliote-

SISTEMA MODUS

cario compie di frequente. Si tratta cioè di intervenire in quel processo di selezione della documentazione che consegna alla storiografia fonti documentarie. L'azione di conservazione e di valorizzazione è il primo passo verso la ricostruzione o costruzione del passato e affianca il lavoro interpretativo della ricerca storiografica.

Certamente quella storiografia che si dedica alla storia dei comportamenti culturali, alla storia del costume e delle mentalità, che indaga la storia dei partiti e dei movimenti politici anche attraverso le biografie e la vita quotidiana delle persone, che studia i miti collettivi, i linguaggi e le emozioni non può che trovare nell'osservazione delle immagini e nell'analisi dei messaggi e dei modelli di comunicazione che i manifesti contengono un materiale di grande interesse. I documenti visivi (manifesti, ma anche foto e audiovisivi) fanno parte, con i carteggi, le autobiografie e i diari, delle "nuove fonti" che negli ultimi decenni hanno particolarmente attirato l'interesse dei ricercatori che si occupano della storia della politica attingendo anche ad altre discipline quali l'antropologia, la sociologia, gli studi culturali.

Manifesto, Per la pace e la rinascita del Mezzogiorno... l'Unità in ogni famiglia, a cura dell'Ufficio Propaganda della Direzione generale de l'Unità Roma, 1951

Il concetto di fonte si è modificato ed esteso. Come scrive Peter Burke in *Testimoni oculari,* «potrebbe essere utile sostituire il concetto di fonti con quello di tracce del passato nel presente. Il termine tracce si riferisce tanto ai manoscritti, ai libri a stampa, agli edifici, agli arredi, al paesaggio (modificato dallo sfruttamento dell'uomo), quanto ai vari tipi di immagine: dipinti, sculture, incisioni, fotografie»[6].

[6] P. Burke, *Testimoni oculari. Il significato storico delle immagini,* Roam, Carocci, 2002.

SISTEMA MODUS

Definire le fonti iconografiche "tracce" rinvia immediatamente ad un altro storico, a Carlo Ginzburg, a partire dal noto saggio dal titolo *Spie. Radici di un paradigma indiziario*[7]. Potrebbe essere questo un modo particolarmente appropriato per pensare alla definizione di fonte storica del documento/manifesto e cioè concepirlo come traccia indiziaria per suggerire, provocare l'indagine storiografica. Lo stesso Carlo Ginzburg ne dà prova nella conferenza alla Biblioteca Nazionale di Francia, tenuta nel 2000, pubblicata in parte nelle pagine culturali di un quotidiano nazionale e in forma integrale su *History Workshop Journal*. L'argomento trattato è l'analisi dell'uso di una stessa immagine, o per meglio dire, dello stesso suggestivo messaggio, nel corso di un lungo periodo storico e in occasione di eventi tra loro diversi. Si tratta del manifesto con il ritratto di Lord Kitchener affisso sui muri di Londra allo scoppio della Prima guerra mondiale. Nel saggio vengono presi in esame contesto storico, efficacia comunicativa, dinamiche psicologiche della comunicazione di massa e storia delle diverse versioni dell'immagine evocatrice di un analogo messaggio. È un esempio molto interessante di come possa venire trattato e studiato un documento con le caratteristiche del manifesto contemporaneo, mettendo in campo categorie interpretative di più discipline.

Oltre al saggio di Carlo Ginzburg anche la mostra di Carlo Quintavalle a Parma nel 1999 dal titolo *Il Rosso e il nero. Figure e ideologie in Italia 1945-1980* molto ha contribuito a far capire il valore del documento - manifesto. Ne rimane testimonianza un voluminoso catalogo dove scrive Quintavalle nel saggio introduttivo:

«Spieghiamo subito perché si deve iniziare dai manifesti e non dall'arte per parlare degli anni del confronto fra sinistra e centro, fra sinistra e destra in Italia …… i manifesti sono l'immagine delle parole, quelle che si sentono alla radio, e dunque per anni sono la sola immagine a colori. Dunque il manifesto è il colore delle idee, è il colore dei personaggi. Ma è anche soprattutto un racconto, una storia, se si vuole una fiaba»[8].

Rinviando alla loro lettura completa, confermo la mia convinzione che i due saggi citati, di Carlo Ginzburg e di Carlo Arturo Quintavalle, siano stati i primi approcci innovativi dell'uso del documento-manifesto come fonte storica, interpretata cogliendo il valore simbolico del colore, del linguaggio allusivo quasi fiabesco, mettendo in campo l'analisi psicologica, il confronto del linguaggio politico con quello della pubblicità, con quello del cinema e dell'arte. Sono stati offerti pertanto nuovi strumenti alla interpretazione del contesto storico del quale sono stati espressione.

Si può capire di conseguenza il grande valore che attribuiamo alla figura di Albe Steiner (1913-1974). Nella banca dati sono visibili 41 manifesti realizzati dal designer e grafico che l'Archivio Albe e Lica Steiner (Politecnico di Milano) - che con-

[7] C. Ginzburg, *Spie. Radici di un paradigma indiziario* in *Miti emblemi spie,* Torino, Einaudi, 1986, p. 158-209.

[8] *Il Rosso e il Nero. Figure e ideologie in Italia 1945-1980 nelle raccolte del CSAC,* a cura di Gloria Bianchino e Arturo Carlo Quintavalle, Milano, Electa, 1999.

SISTEMA MODUS

Manifesto della Comissao pastoral da Terra del Brasile, disegnato da Elifas Andreato [S.l., 1995]

serva tutta la documentazione relativa all'attività personale e professionale di Albe Steiner nei settori della grafica e del disegno industriale - ha voluto rendere consultabili attraverso questa risorsa in rete che è manifestipolitici.it. Sono documenti di propaganda politica e sociale elaborati da Albe Steiner tra il 1948 e il 1973. L'insieme comprende manifesti per il Partito comunista Italiano (sottoscrizioni e Feste de l'Unità), per la diffusione del quotidiano l'Unità, per l'Associazione Nazionale Ex Deportati politici nei campi di sterminio nazisti, molti realizzati anche con la partecipazione degli studenti del Convitto Rinascita, dell'Umanitaria, del Movimento studentesco. Tra i temi ricorrenti la deportazione nazista, la lotta per la pace e contro la bomba atomica, la festa della Liberazione del 25 aprile, la festa del Primo maggio, la solidarietà al popolo cileno, la rivoluzione cubana.

Un altro aspetto che vorrei considerare alla fine di queste pagine è il grande contributo dato al trattamento e alla valorizzazione di questo genere di documenti dall'informatica e dal Web, permettendone la consultazione ad un pubblico ampio e sottraendo il documento originale all'usura e alla manipolazione.

Per la descrizione del documento si è scelto l'uso del software SEBINA aggiornato progressivamente nelle sue varie versioni perché permetteva un'ampia descrizione del documento singolo secondo standard bibliografici internazionali. Si è cercato di offrire alla ricerca il maggior numero possibile di dati non solo sul documento singolo ma anche sul contesto documentario o sulla descrizione del soggetto produttore o conservatore. Chi consulta la banca dati trova dunque oltre alla riprodu-

SISTEMA MODUS

zione digitale del manifesto la descrizione del contenuto secondo gli standard ISBD non book. Per autore si è considerato il grafico, il committente, il fotografo, o l'autore di una citazione presente nel testo. Sono descritti anche i soggetti del contenuto dei documenti, comprese le immagini. Chiunque consulta ha pertanto a disposizione uno spettro piuttosto ampio di informazioni e di accessi.

Chi sono coloro che consultano la nostra banca dati? Da questa esperienza si conferma quanto tanti hanno scritto: Internet promuove la funzione sociale dei patrimoni documentari. Il mondo degli utenti è esteso e composito: oltre a studiosi e ricercatori c'è un numero consistente di case editrici, di testate di periodici a contenuto storico, di istituti culturali (italiani e non), di enti pubblici e privati promotori di mostre, di scuole o di singoli insegnanti per scopi didattici. Non mancano privati cittadini che desiderano riproduzioni per ragioni affettive; giungono richieste per set teatrali, televisivi, cinematografici che hanno bisogno di manifesti per la ricostruzione storica ambientale.

Concludo con una citazione di Giovanni De Luna che in sintesi riassume con semplicità quella che è la storia del manifesto politico. È riferita ad un evento storico preciso, ma può valere per molti altri

«Quando tutto finì, i manifesti furono stracciati, i muri ridipinti, le cartacce spazzate via dalle strade del Quartiere Latino finalmente purificate. Ma ci fu qualcuno che li tenne come ricordo, altri che ne depositarono qualche copia negli archivi quasi presagendo quello che sarebbero diventati in futuro, quello che sono diventati oggi , tracce , appunto, documenti che ci consentono – a distanza di decenni – di avvicinarci alla realtà profonda di quell'evento»[9].

[9] Giovanni De Luna, *La politica sui muri, i manifesti del '68*, in *L'arte per la strada. I manifesti del maggio francese,* Giulio Bolaffi editore, Torino 2008, p.47-58.

L'ultima consultazione dei siti Web è avvenuta nel mese di dicembre 2015.

SISTEMA MODUS

Eventi

Beni culturali ecclesiastici e comunicazione. *Un umanesimo digitale possibile?*

Presentazione del Portale BeWeb completo di beni storici e artistici, architettonici, archivistici, librari e di descrizioni di istituti culturali di conservazione

Roma, 12-13 maggio 2015

Don Valerio Pennasso
Direttore dell'Ufficio Nazionale per i beni culturali della CEI

Si è tenuta a Roma il 12 e il 13 maggio scorso la XXII Giornata Nazionale promossa dall'Ufficio Nazionale per i beni culturali ecclesiastici della Conferenza Episcopale Italiana sul tema *Beni culturali ecclesiastici e comunicazione. Un umanesimo digitale possibile?* articolato intorno all'analisi del metodo di lavoro, dei risultati raggiunti e dei prossimi obiettivi legati al Portale BeWeb[1] .

L'evento ha coinvolto 219 partecipanti, in rappresentanza di 111 Diocesi italiane e di tutte le 16 Regioni Ecclesiastiche, oltre a numerosi istituti culturali e rappresentati del MiBact e degli Istituti Centrali dello Stato.

Sul palco si sono alternati relatori di diversa formazione ed esperienza umana e professionale, con l'obiettivo di mettere a confronto metodi ed approcci operativi diversi, che hanno trovato terreno di dialogo nel cosiddetto "*umanesimo digitale*".

Il *focus* della riflessione è stato il Portale BeWeb, luogo virtuale di consultazione di descrizioni di beni storici e artistici, architettonici, archivistici, librari e degli istituti di conservazione che raccoglie ed espone il lavoro di censimento del patrimonio che la Chiesa che è in Italia sta realizzando per i beni culturali di sua proprietà ed un primo set significativo di servizi rivolti sia alle realtà ecclesiali che al mondo del Web.

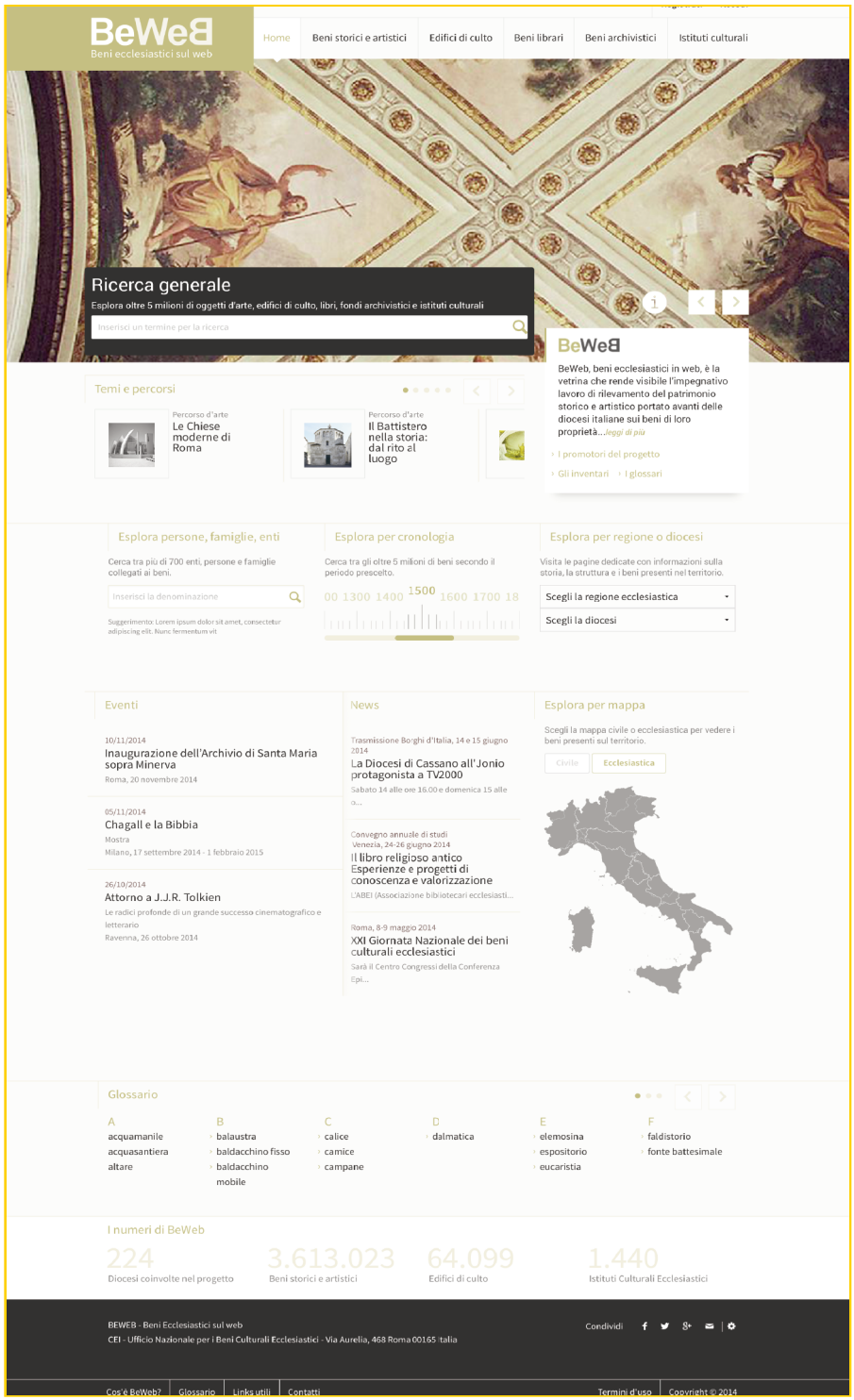
La messa a disposizione del portale trasversale BeWeb è solo un primo passo, incentrato su un'ampia possibilità di ricerca e di correlazione tra i dati.

BeWeb oggi ospita oltre 3.800.000 beni storici e artistici inventariati; oltre 65.500 edifici di culto segnalati dalle diocesi di cui 13.300 beni architettonici censiti; oltre 500.000 record bibliografici; 149 banche dati di archivi storici – oltre 3.300 fondi e complessi di fondi; oltre 1.100 Authority File (Persona e famiglia, gli Enti saranno caricati successivamente); 1.483 istituti descritti (archivi, biblioteche, musei). BeWeb verrà periodicamente implementato con nuovi dati.

Hanno aperto i lavori della prima giornata gli interventi di S.E. Mons. Nunzio Galantino, Segretario Generale della CEI e dell'architetto Antonia Pasqua Recchia, Segretario Generale del MiBact, che ribadiscono il clima di piena e fattiva collaborazione esistente tra le rispettive istituzioni in attuazione delle diverse Intese

[1] www.chiesacattolica.it/beweb.

SISTEMA MODUS


BeWeB
Beni ecclesiastici sul web
Home
Beni storici e artistici
Edifici di culto
Beni librari
Beni archivistici
Istituti culturali
Ricerca generale
Esplora oltre 5 milioni di oggetti d'arte, edifici di culto, libri, fondi archivistici e istituti culturali
Inserisci un termine per la ricerca
BeWeB
BeWeB, beni ecclesiastici in web, è la vetrina che rende visibile l'impegnativo lavoro di rilevamento del patrimonio storico e artistico portato avanti delle diocesi italiane sui beni di loro proprietà...leggi di più
I promotori del progetto
Gli inventari
I glossari
Temi e percorsi
Percorso d'arte
Le Chiese moderne di Roma
Percorso d'arte
Il Battistero nella storia: dal rito al luogo
Esplora persone, famiglie, enti
Cerca tra più di 700 enti, persone e famiglie collegati ai beni.
Inserisci la denominazione
Esplora per cronologia
Cerca tra gli oltre 5 milioni di beni secondo il periodo prescelto.
00 1300 1400 1500 1600 1700 18
Esplora per regione o diocesi
Visita le pagine dedicate con informazioni sulla storia, la struttura e i beni presenti nel territorio.
Scegli la regione ecclesiastica
Scegli la diocesi
Eventi
10/11/2014
Inaugurazione dell'Archivio di Santa Maria sopra Minerva
Roma, 20 novembre 2014
05/11/2014
Chagall e la Bibbia
Mostra
Milano, 17 settembre 2014 - 1 febbraio 2015
26/10/2014
Attorno a J.J.R. Tolkien
Le radici profonde di un grande successo cinematografico e letterario
Ravenna, 26 ottobre 2014
News
Trasmissione Borghi d'Italia, 14 e 15 giugno 2014
La Diocesi di Cassano all'Jonio protagonista a TV2000
Sabato 14 alle ore 16.00 e domenica 15 alle o...
Convegno annuale di studi
Venezia, 24-26 giugno 2014
Il libro religioso antico Esperienze e progetti di conoscenza e valorizzazione
L'ABEI (Associazione bibliotecari ecclesiasti...
Roma, 8-9 maggio 2014
XXI Giornata Nazionale dei beni culturali ecclesiastici
Sarà il Centro Congressi della Conferenza Epi...
Esplora per mappa
Scegli la mappa civile o ecclesiastica per vedere i beni presenti sul territorio.
Civile
Ecclesiastica
Glossario
A
acquamanile
acquasantiera
altare
B
balaustra
baldacchino fisso
baldacchino mobile
C
calice
camice
campane
D
dalmatica
E
elemosina
espositorio
eucaristia
F
faldistorio
fonte battesimale
I numeri di BeWeb
224
Diocesi coinvolte nel progetto
3.613.023
Beni storici e artistici
64.099
Edifici di culto
1.440
Istituti Culturali Ecclesiastici
BEWEB - Beni Ecclesiastici sul web
CEI - Ufficio Nazionale per i Beni Culturali Ecclesiastici - Via Aurelia, 468 Roma 00165 Italia
Condividi
Cos'è BeWeb?
Glossario
Links utili
Contatti
Termini d'uso
Copyright © 2014

ed Accordi intercorsi a partire dal 1996 (Intesa 1996, Intesa 2000, Intesa 2005; Accordi specifici).

Monsignor Stefano Russo, direttore dell'Ufficio Nazionale per i beni culturali ecclesiastici della CEI, delinea le caratteristiche fondanti del servizio. L'impegno dell'Ufficio in questi anni si è concretizzato nel coordinamento di un lavoro, incentrato sulla qualità, che ha prediletto il continuo confronto con le Chiese particolari (225 Diocesi in Italia) e le istituzioni dello Stato.

Paul Gabriele Weston, professore all'Università degli Studi di Pavia e referente scientifico dei progetti CEI dedicati ad archivi e biblioteche, traccia i contorni di un progetto culturale che non ha mai perso di vista il confronto tra discipline e percorsi differenti che hanno trovato felice sintesi nel Portale BeWeb. Il lavoro è concepito per il dialogo con altri sistemi informativi, nazionali e internazionali, ed è in continua evoluzione. Se questa prima fase ha privilegiato la pura consultazione, su base dati disciplinare o fra più discipline, i prossimi passi andranno nella direzione dei servizi e della corretta valorizzazione pastorale.

A Weston sono seguiti tre interventi che descrivono le esperienze più recenti di proficua collaborazione tra CEI e MiBact (Prefetto Fabio Carapezza Guttuso, Tenente Colonnello Roberto Colasanti, Maria Idria Gurgo in sostituzione della Giannetto)

Pier Cesare Rivoltella e Gino Roncaglia suggeriscono lo sviluppo nel Portale BeWeb di servizi fruibili in mobilità e a ulteriore utilità di un utente sempre più esigente.

Conclude la giornata l'intervento di Benedetto Ponti che illustra le caratteristiche giuridiche delle banche dati messe a disposizione su Web.

La seconda giornata si apre con l'intervento di Antonio Paolucci che ha illustrato il valore del patrimonio italiano diffuso sul territorio sottolineando la cura che esso richiede.

Gli interventi successivi hanno natura strettamente tecnica e focalizzano l'attenzione sul percorso di progettazione e realizzazione del portale BeWeb. Federico Parrella illustra come per il *redesign* di BeWeb (su Web fin dal 2000) si è adottato l'approccio *User Centered Design* e il modello *win/win*, prevedendo l'individuazione di *personas* (utenti tipo attesi e ricercati) e delineando scenari d'uso specifici per il progetto. Illustra, infine, la responsività del sito per ottenere l'usabilità di BeWeb in nuovi scenari e contesti d'uso. Leo Spadaro illustra invece le APP in corso di realizzazione per la fruizione dei beni culturali ecclesiastici: le Cattedrali d'Italia; le sperimentazioni per visite a distanza e immersive (Bergamo, Catania, Firenze); l'anagrafe degli istituti culturali (già scaricabile gratuitamente).

In conclusione, BeWeb da un senso pieno all'umanesimo digitale perché è un servizio crossmediale come estensione della caratteristica dell'uomo di fare sintesi fra diverse competenze e conoscenze e ancora perché, in quanto banca dati, costituisce un *servizio per la memoria*, base per la costruzione del futuro dell'uomo.

SISTEMA MODUS

Convegno Anno Europeo 2015 "Gli Obiettivi di Sviluppo sostenibile dell'Onu e la Carta di Milano"

Roma, 8 giugno 2015

Massimina Cattari
ICCU - CDE OPIB

Nell' ambito dell'Anno Europeo 2015[1] si è tenuto a Roma, lunedì 8 giugno 2015 il Convegno dal titolo "Gli Obiettivi di Sviluppo sostenibile dell'Onu e la Carta di Milano"[2] presso la prestigiosa sede della Società Italiana per l'Organizzazione Internazionale (SIOI).

Il convegno è stato ideato e organizzato congiuntamente dai 5 Centri di Documentazione Europea (CDE) di Roma di: CNR, SIOI, ICCU-OPIB, "Sapienza" Università di Roma, SNA con la collaborazione e il finanziamento della Rappresentanza in Italia della Commissione Europea[3].

L'iniziativa si inserisce nell'ambito del Progetto nazionale 2015 Anno Europeo per lo Sviluppo 2015 *"Il nostro mondo, la nostra dignità, il nostro futuro"*[4] della Rete italiana dei CDE e della Rappresentanza in Italia della Commissione Europea.

A poco più di un mese dall'apertura di Expo 2015 (1°maggio - 31 ottobre 2015)[5], lo scopo del convegno è stato quello di porre al centro dell'attenzione temi di grande attualità: la partecipazione dell'Unione Europea a Expò Milano 2015; l'approfondimento della *Carta di Milano*; il tema dei "*Nuovi Obiettivi di Sviluppo Sostenibile*".

Ricercatori ed esperti provenienti da vari ambiti istituzionali, con i loro interventi, hanno af-

[1] Il sito Web dell' Anno Europeo dello Sviluppo 2015: <https://europa.eu/eyd2015/>. L'Anno europeo è una delle principali Iniziative UE e rappresenta un'importante strumento di informazione e promozione dell'UE. La Decisione (UE) n. 472/2014 del PE e del Consiglio, 16/04/2014, istitutiva dell'Anno Europeo dello Sviluppo (2015) è stata pubblicata sulla GUCE/GUUE L del 09/05/2014 Serie L: <http://eur-lex.europa.eu/legal-content/IT/TXT/?uri=uriserv:OJ.L_.2014.136.01.0001.01.ITA>. Per informazioni più dettagliate consultare la Sezione relativa del sito OPIB:<http://www.opib.librari.beniculturali.it/index.php?it/464/2015-anno-europeo-dello-sviluppo>.

[2] Il Programma del Convegno: <http://www.opib.librari.beniculturali.it/index.php?it/145/eventi/138/convegno-gli-obiettivi-di-sviluppo-sostenibile-dellonu-e-la-carta-di-milano>.

[3] Il Sito Web della Rappresentanza in Italia della Commissione Europea:<http://ec.europa.eu/italy/index_it.htm>.

[4] Il Sito Web della Rete italiana dei Centri di Documentazione Europea – CDE: <http://www.cdeita.it/>.

[5] Il sito ufficiale di ExpoMilano2015 <http://www.expo2015.org/it/index.html>. *Expo Milano 2015* è l'Esposizione Universale che l'Italia ospiterà dal *1° maggio al 31 ottobre 2015* e sarà il più grande evento mai realizzato sull'alimentazione e la nutrizione. Per sei mesi Milano diventerà una vetrina mondiale in cui i Paesi mostreranno il meglio delle proprie tecnologie per dare una risposta concreta a un'esigenza vitale. *Expo Milano 2015* sarà la piattaforma di un confronto di idee e soluzioni condivise sul tema dell'alimentazione, stimolerà la creatività dei Paesi e promuoverà le innovazioni per un futuro sostenibile.

SISTEMA MODUS

Figura 1: *La locandina del Convegno "Gli obiettivi di sviluppo sostenibile dell'ONU e la carta di Milano. SIOI - Roma 8 giugno 2015*

frontato i temi oggetto del convegno e dato il loro contributo per comporre un quadro generale delle iniziative attuate in Europa e in Italia.

Il convegno è iniziato con i saluti delle istituzioni che ospitano i CDE ed in particolare ha aperto i lavori il Presidente della SIOI Franco Frattini, seguito da Giovanni Tria, Presidente della Scuola Nazionale dell'Amministrazione di Roma (SNA), Simonetta Buttò, Direttore dell'Istituto centrale per il catalogo unico delle bBiblioteche italiane e per le informazioni bibliografiche (ICCU), Roberto Volpe in sostituzione di Paolo Annunziato Direttore Generale del CNR e Umberto Triulzi, Delegato del Rettore per la Cooperazione internazionale della "Sapienza" Università di Roma.

Il moderatore del convegno è stato Vittorio Calaprice, Responsabile dei Rapporti con le istituzioni nazionali e gli stakeholders della Rappresentanza in Italia della Commissione Europea.

La relatrice Rossella Speroni (Joint Research Center- JRC della Commissione Europea), con il suo intervento *"La partecipazione e il contributo dell'Unione Europea a EXPO Milano 2015"* ha presentato il contesto dell'Expo 2015 di Milano e la partecipazione delle organizzazioni internazionali tra cui l'Unione Europea[6] (presente con un proprio padiglione *EU Pavilion)* a questo importante evento. R. Speroni, cita il motto della partecipazione dell'UE *"Coltivare insieme il futuro dell'Europa per un mondo migliore"* che ha come primo obiettivo quello di comunicare un importante messaggio di valori ai visitatori: l'impegno dell'UE per la creazione di un'Europa e di un mondo sempre più sostenibile.

Le *Organizzazioni Internazionali*, insieme a

[6] Il sito Web del Padiglione UE: <http://europa.eu/expo2015/it/>.
Per informazioni più dettagliate consultare la Sezione relativa del sito OPIB:
<http://www.opib.librari.beniculturali.it/index.php?it/486/expo-milano-2015-e-l-unione-europea>.

Expo Milano 2015 e agli altri partecipanti, svilupperanno attività ed eventi (artistici e musicali, convegni, spettacoli, laboratori creativi e mostre) durante l'Esposizione per portare all'attenzione dei visitatori la propria expertise e le proprie best practices.

Inoltre, la relatrice ha parlato del Centro Comune di Ricerca – Joint Research Centre - JRC[7] e del Servizio Volontari[8] per il Padiglione Ue, organizzato e gestito dallo stesso JRC, ha fornito ai giovani partecipanti al convegno dettagliate informazioni della call e delle modalità di partecipazione, sottolineando che per i volontari sarà una grande esperienza formativa e di comunicazione.

Vittorio Calaprice (Responsabile dei Rapporti con le istituzioni nazionali e gli stakeholders della Rappresentanza in Italia della Commissione Europea), in sostituzione dell'intervento di Antonio Bettanini**,** ha presentato "*La Carta di Milano*"[9], il documento che costituirà l'eredità di *Expo Milano 2015* rivolto a cittadini, istituzioni, imprese, associazioni e ai Paesi che vorranno sottoscriverla. Lo stesso documento, può essere definito come una sorta di sintesi nata dalle proposte e dalle idee discusse durante l'evento "Expo delle Idee" svoltosi a Milano il 7 febbraio 2015, dove 500 esperti organizzati in più di quaranta tavoli tematici si sono confrontati sullo sviluppo sostenibile e sul diritto al cibo. Il secondo appuntamento su "*La Carta di Milano*" si è svolto a Firenze il 28 marzo 2015, nell'evento dal titolo "Italia 2015". Il Paese nell'anno dell'Expo" a cui ha partecipato il Presidente della Repubblica italiana Sergio Mattarella. Un terzo incontro è stato l'evento svoltosi a Pompei il 18 aprile 2015 dove il Ministro delle politiche agricole alimentari e forestali Maurizio Martina ha presentato i principali appuntamenti che si svolgeranno durante il semestre all'interno del Sito Espositivo.

Il protocollo per la definizione dei "*Nuovi Obiettivi del Millennio*"[10] scritta per tutti i Paesi che decideranno di aderirvi, è stata consegnata a *Ban Ki-Moon*, l'attuale Segretario Generale delle Nazioni Unite, il 16 ottobre 2015 in occasione della "*Giornata Mondiale dell'Alimentazione*"[11] per definire l'atto d'impegno che il Governo italiano proporrà al mondo come lascito dell'Esposizione Universale.

Paolo Bargiacchi (Ordinario di diritto internazionale, KORE Enna) nel suo intervento" *L'Agenda Post 2015 sullo sviluppo sostenibile*" ha illustrato gli obiettivi dell'*Agenda for Change (2014-2020)*[12] e ha approfondito i temi legati alle politiche pubbliche europee, alla crescita e allo sviluppo, con riferimento sia al contesto europeo sia al ruolo svolto dall'Europa nei confronti dei Paesi extra europei meno economicamente avanzati. In merito al contributo dell'Europa sui temi dell'alimentazione l'UE tra i suoi obiettivi ha quello di debellare la fame nel mondo grazie alla sua azione di cooperazione allo sviluppo. Il prolungato periodo di crisi che ha investito molte economie mondiali, costituisce infatti un momento fondamentale per riflettere sulla co-

[6] Il sito Web del Padiglione UE: <http://europa.eu/expo2015/it/>.

[7] Il sito Web del Centro Comune di Ricerca – Joint Research Centre - JRC <https://ec.europa.eu/jrc/>.

[8] Il sito Web del Servizio Volontari UE <http://www.expo2015.volunteer4europe.org/> per effettuare la registrazione on-line per candidarsi durante i sei mesi di *EXPO Milano 2015*.

[9] *"La Carta di Milano" <http://www.expo2015.org/it/la-carta-di-milano>. Per approfondimenti visita il sito dedicato: <http://carta.milano.it/it/>.*

[10] I nuovi obiettivi del Millennio:< http://www.unric.org/it/global-action-2015>.
Il sito Web dell'ONU: <http://www.un.org/en/index.html>.

[11] Per maggiori informazioni visita la pagina ufficiale della Giornata Mondiale dell'Alimentazione: < http://www.fao.org/world-food-day/home/it/>

[12] *"Agenda for Change (2014-2020)" : European Commission- Brussels, 13.10.2011 Doc.COM (2011) 637 final: <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX%3A52011DC0637&qid=1412922281378&from=EN>.*

SISTEMA MODUS

Figura 2: *Saluti istituzionali.*
Al centro Franco Frattini Presidente SIOI, da sinistra Simonetta Buttò Direttore ICCU,Giovanni Tria Presidente SNA, Roberto Volpe CNR,Umberto Triulzi Delegato del Rettore per la Cooperazione Internazionale – Università "Sapienza"

Figura 3: *Partecipanti al Convegno.*
Al centro Sara Cavelli Direttore Generale e Responsabile CDE della SIOI

Figura 4: *Moderatore e Relatori del convegno*
Al centro Vittorio Calaprice Responsabile dei Rapporti con le istituzioni nazionali e gli stakeholders della Rappresentanza in Italia della Commissione Europea, a sinistra Paolo Bargiacchi Ordinario di Diritto Internazionale, Kore Enna, a destra Rossella Speroni Joint Research Center – JRC della Commissione Europea

Figura 5: *Simonetta Buttò Direttore dell'ICCU*

Figura 6: *Responsabili della Rete dei Centri di Documentazione Europea di Roma e promotori del Convegno. Da sinistra Franco Botta CDE Università "Sapienza",Maria Laura Viora CDE SNA, Massimina Cattari CDE – OPIB ICCU, Renato Genovese CDE SIOI, M. Adelaide Ranchino CDE CNR*

Figura 7: *Giovane partecipante e consegna dei materiali informativi convegno della Commissione Europea*

struzione europea e sulla sua capacità di azione e di sviluppo sia all'interno sia nei confronti dei Paesi terzi. Infine, il relatore ha illustrato le attività svolte dalla DG International Cooperation and Development[13].

[13] Il sito Web della DG International Cooperation and Development: <http://ec.europa.eu/europeaid/index_en.htm>.

A chiusura del Convegno è intervenuta Giada Dionisi (SIOI-MSOI Movimento Studentesco per l'Organizzazione Internazionale) ha presentato "ROMUN 2015" (Rome Model United Nations), la simulazione giovanile internazionale *Model UN* che si è tenuta recentemente a Roma dal 16 al 20 ottobre 2015. La relatrice, inoltre, ha fornito ai giovani partecipanti al convegno dettagliate informazioni in merito alla partecipazione dell'evento mediante iscrizione sui siti dedicati[14].

Al convegno hanno partecipato più di 150 persone tra studenti universitari, docenti, ricercatori, bibliotecari, giornalisti, funzionari della PA e cittadini interessati alle tematiche del convegno. Hanno partecipato, inoltre, gli studenti iscritti ai master della SIOI. L'evento è stato molto apprezzato da parte dei partecipanti (considerati i temi di grande interesse e attualità trattati durante il convegno) i quali hanno auspicato una maggiore frequenza di iniziative di questo genere.

La realizzazione di questa iniziativa ha permesso ai 5 CDE di Roma di consolidare ulteriormente i rapporti con le altre istituzioni europee e nazionali. Tutti i relatori sono stati molto disponibili con i partecipanti al convegno ed hanno risposto a tutte le loro domande contribuendo al successo dell'iniziativa.

Grazie al finanziamento della Rappresentanza in Italia della Commissione Europea si è potuto realizzare sia le borse con i loghi delle istituzioni dei 5 CDE di Roma, sia una brochure informativa, aggiornata, contenente le informazioni sulla Rete dei CDE in Italia e i riferimenti dei 5 CDE romani.

Inoltre, all'ingresso del Salone della SIOI durante il Convegno sono stati allestiti dei desk informativi con documentazione preparata dai 5 CDE che hanno promosso e organizzato l'evento e messo a disposizione dei partecipanti anche le varie brochure dei CDE e altro materiale informativo delle istituzioni di appartenenza oltre a quello fornito dai vari enti che hanno sostenuto l'iniziativa.

L'evento è stato pubblicizzato non solo sui siti dei singoli CDE ma anche sui propri siti istituzionali di SIOI, CNR, ICCU, "Sapienza", SNA oltre ai Social Network, liste di discussione, istituzioni culturali e di ricerca, newsletter, riviste professionali specializzate, comunicati stampa, mailing list dei CDE. La promozione del convegno si è realizzata anche con l'affissione di locandine in tutte le sedi delle istituzioni dei CDE qualche settimana prima dell'evento.

Infine, il Convegno*: "Gli Obiettivi di Sviluppo sostenibile dell'Onu e la Carta di Milano"* ha avuto una particolare rilevanza anche perché si è inserito tra i principali eventi nazionali ed europei[15] volti a sensibilizzare l'opinione pubblica e i cittadini europei sulle tematiche dell'Anno Europeo per lo Sviluppo 2015, designato come tale dall'Unione Europea.

[14] I siti Web: <http://www.sioi.org> e < http://www.romemun.org/>.

[15] Il Convegno è stato pubblicato sul Portale europeo dell'Anno europeo per lo Sviluppo 2015 nella Sezione Eventi: https://europa.eu/eyd2015/it/europe-direct/events/onu

L'ultima consultazione dei siti Web è avvenuta nel mese di dicembre 2015.

SISTEMA MODUS

Segnalazioni

Il progetto Europeana Food and Drink

Il cibo, il bere e il patrimonio culturale

Elisa Sciotti
ICCU

Il progetto Europeana Food and Drink[1] esplora i diversi aspetti della cultura alimentare mettendo in risalto il profilo domestico, quello sociale e quello maggiormente connesso all'industria e alla distribuzione. Il progetto sostiene il riuso dei contenuti culturali digitali che riguardano il cibo e le bevande da parte dell'industria creativa affinché quest'ultima possa trarne beneficio dal punto di vista sia economico sia, appunto, creativo.

Europeana Food and Drink, cui partecipano 28 istituzioni di tutta Europa, è cofinanziato dalla Commissione Europea ed è coordinato dall'associazione Collections Trust che ha sede a Londra; lanciato nel gennaio 2014, Food and Drink si concluderà nel maggio 2016.

L'ICCU partecipa all'iniziativa attraverso i contenuti forniti dalla Biblioteca Universitaria Alessandrina, dalla Biblioteca Nazionale Centrale di Roma e dalla Soprintendenza per i beni archeologici del Lazio e dell'Etruria meridionale.

Il progetto Food and Drink è incluso nel gruppo di progetti collegati a Europeana[2], il grande portale sul patrimonio culturale dove si possono trovare milioni di contenuti tra testi, immagini, filmati, documenti sonori e oggetti 3D, provenienti da migliaia di istituzioni europee. Attualmente il portale rende consultabili oltre 48.000.000 di contenuti che aumenteranno nel prossimo futuro.

Ci si attendono molteplici risultati da Food and Drink. Oltre all'area del portale Europeana con più di 50.000 nuovi contenuti digita-

Mostra virtuale "Cake? Cake!"

[1] http://foodanddrinkeurope.eu/.
[2] http://www.europeana.eu/portal/.

Logo di Europeana Food and Drink

li[3], sono previsti per l'intera durata del progetto eventi per incoraggiare il riuso dei contenuti culturali legati al mondo dell'alimentazione al fine di coinvolgere e interessare il grande pubblico, i rivenditori e i distributori. Fra gli eventi organizzati vi sono, ad esempio, gli *Open Innovation Challenges* che consentono di sviluppare e consolidare la cooperazione con partner dell'industria creativa per il riutilizzo del patrimonio culturale digitale. Il terzo Innovation Challenge[4], realizzato attraverso la cooperazione tra l'Instituto Andaluz del Patrimonio Historico (IAPH) e DigiLab dell'Università la Sapienza di Roma, si sta svolgendo nei giorni in cui viene redatto questo articolo e terminerà il 20 dicembre: i partecipanti sono invitati a presentare un prodotto – si può scegliere fra la categoria 2D o 3D– realizzato attraverso il riuso di contenuti culturali presenti in Europeana e relativi al progetto Food and Drink. Questa competizione mira, appunto, a creare un legame tra il patrimonio digitale culturale e l'industria creativa e a valorizzare cibi e saperi locali. Al vincitore di ciascuna categoria, nel corso di una cerimonia che si svolgerà a Siviglia il 29 gennaio 2016, verrà assegnato un premio di 2.000 euro.

La cultura del cibo, in questo caso del bere, passa anche per le strade di Londra, attraverso la piattaforma *Tea Trail London*[5], realizzata nell'ambito del progetto Europeana Food and Drink dall'Horniman Museum and Gardens e da Semantika. *Tea Trail London* consente di compiere, attraverso tre percorsi, un coinvolgente viaggio nella Londra del tè fra storia, sapori, tradizioni, sale dedicate, collezioni museali e botteghe.

E dopo aver vagato nelle strade del tè, ci si può concedere un altro tipo di divertimento con *Food Planet*[6], il social game sviluppato dalla compagnia AcrossLimits e l'International Centre for Information Management Systems and Services per Food and Drink. I partecipanti devono indovinare il paese d'origine di

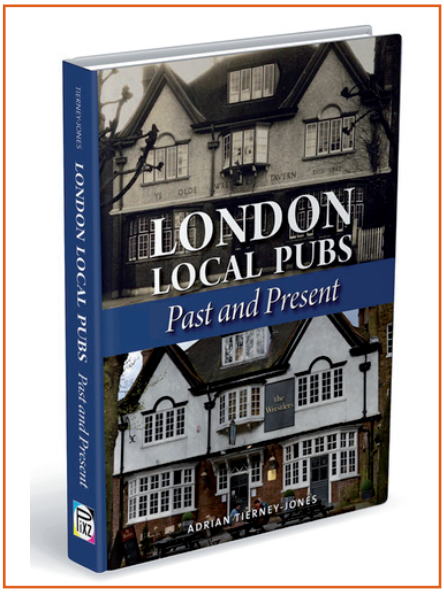
[3] http://www.europeana.eu/portal/search?/[provider][]=europeana+food+and+drink.

[4] Maggiori informazioni sulla competizione sono disponibili nella sezione dedicata del sito Web di Europeana Food and Drink, <http://foodanddrinkeurope.eu/get-involved/challenges-and-events/challengethree/>.

[5] <http://teatrail.london/>. Per un ulteriore introduzione alla piattaforma si rimanda alla pagina dedicata nel sito Europeana Food and Drink, <http://foodanddrinkeurope.eu/applications/food-trails/>.

[6] http://foodanddrinkeurope.eu/applications/games.

SISTEMA MODUS



Libro London Local Pubs: Past and Present di Adrian Tierney-Jones

piatti, bevande, strumenti legati al mondo dell'alimentazione e ingredienti in un coinvolgente gioco. Oltre ad apprendere e scoprire, i giocatori potranno anche intervenire condividendo le proprie ricette predilette. *Food Planet*[7], attualmente disponibile per Android, è raggiungibile anche attraverso la pagina Facebook dedicata.

Per tornare nuovamente ai musei, Food and Drink ha allestito una mostra virtuale e una itinerante dedicata alle torte, grande classico della tradizione pasticcera.

La mostra virtuale *Cake!* [8] consente di viaggiare fra diversi paesi d'Europa e anche nel tempo, leggendo notizie e scorrendo immagini. Si narra delle torte tradizionali, del loro legame a periodi dell'anno, a ricorrenze e memorie. Alla mostra virtuale se ne affianca una reale composta di pannelli illustrativi in itinere in varie città del Belgio[9]. Le mostre, qui solo brevemente illustrate, sono state curate dal Centre for Agrarian History e dal Royal Museums of Art and History.

Europeana Food and Drink ha declinato il tema legato al bere e al cibo anche in campo editoriale: nell'ambito del progetto, infatti, è stato realizzato il volume *London Local Pubs: Past and Present*[10]. Scritto da Adrian Tierney-Jones, il libro, realizzato con la collaborazione dei partner HistoryPin e Alinari, regala una veduta sui pub offrendo immagini inedite e testi volti a raccontare il ruolo che questi locali hanno sempre rivestito nella società nel corso degli anni. Il libro è disponibile nel formato classico e e-book.

Il tour di questi celeberrimi locali non deve restare necessariamente nelle pagine del libro, ma può continuare utilizzando HistoryPin[11], la piattaforma attraverso la quale si può partecipare attivamente fornendo anche contributi personali e continuando a scoprire e raccontare, in questo caso, i pub inglesi.

Presto sarà disponibile anche l'e-book *Tasting*

[7] Il gioco è disponibile per Android, <https://play.google.com/store/apps/details?id=com.acrosslimits.foodplanet> oppure è raggiungibile alla pagine dedicate su Facebook, <https://www.facebook.com/foodplanetgame> o sul sito Web del Progetto, <http://foodanddrinkeurope.eu/applications/games/social-game-2/>.

[8] <https://www.google.com/culturalinstitute/exhibit/cake-cake/QRtG6e>. La mostra virtuale Cake! può essere consultata anche attraverso Europeana exhibition, < http://exhibitions.europeana.eu/> e Google Cultura Institute, <https://www.google.com/culturalinstitute/exhibit/cake-cake/QRtG6eMd>.

[9] Per maggiori informazioni su date e luoghi si veda la pagina dedicata, <http://foodanddrinkeurope.eu/applications/cake-exhibition/touring-exhibition/>.

[10] Pagina dedicata nel sito Web del progetto, < http://foodanddrinkeurope.eu/applications/book/london-local-pubs-past-and-present/>. Per maggiori informazioni sui contenuti, <http://unbouncepages.com/londonlocalpubs/>.

[11] <http://www.historypin.org/en/a-history-of-pubs/>. Historypin è una piattaforma realizzata da Shift; per maggiori informazioni, <http://www.shiftdesign.org.uk/products/historypin/>.

SISTEMA MODUS

Welfare. For one penny children at Holy Trinity (L.C.C.) school, Lambeth, are provided with a bottle of warmer milk, which they drink by the hygienic use of straws. 31 October 1929. Crediti: TopFoto.co.uk / EUFD

Historic Europe – Exploring the culinary threads between Austria and Lithuania[12] realizzato nell'ambito del progetto dalla Vilnius University – Faculty of Communication e dall'Austrian National Library. L'opera esplora attraverso le ricette nazionali, le connessioni che intercorrono tra la cucina austriaca e quella lituana.

Visto che le rappresentazioni legate al mondo del cibo e del bere sono generalmente molto interessanti e accattivanti, Europeana Food and Drink, con la collaborazione del partner Topham Partners LLP (TopFoto), ha realizzato una *Picture Library* che contiene immagini riferite al tema del cibo e del bere, consultabili e in vendita online[13].

Non mancano nell'ambito di Food and Drink iniziative di carattere educativo[143] come, ad esempio, le lezioni relative al mondo dell'alimentazione: composte da documenti e strumenti didattici, sono messe a disposizione degli insegnanti e sono consultabili online sul sito del progetto. Le lezioni sono state realizzate grazie al lavoro svolto dal Cyprus Food Museum, dal DigiLab dell'Università la Sapienza di Roma (con il contributo del Museo del Cioccolato Antica Norba) e dell'Irish Local Government Management Agency[15].

[12] http://foodanddrinkeurope.eu/applications/book/tasting-historic-europe-exploring-the-culinary-threads-between-austria-and-lithuania/.

[13] http://eufoodanddrink.eu. Per ulteriori informazioni, <http://foodanddrinkeurope.eu/applications/picture-library>.

[14] http://foodanddrinkeurope.eu/applications/elearning-resources/.

[15] http://foodanddrinkeurope.eu/applications/elearning-resources/lesson-plans/.

SISTEMA MODUS

Nell'ambito delle risorse dedicate all'educazione il DigiLab dell'Università la Sapienza di Roma ha sviluppato un importante strumento chiamato *Europeana attribution tools*[16] che consente di citare la fonte in modo autorevole, semplice e veloce quando vengono utilizzate risorse tratte dal portale Europeana.

Food and Drink è un progetto quanto mai attivo, non solo nel suo importante obiettivo di fornire considerevoli contenuti al portale Europeana, ma anche nel creare una rete di cooperazione e di attività per promuovere il riuso e la valorizzazione dei contenuti culturali. Le sue molteplici attività mirano coinvolgere il pubblico e invitare nuove Istituzioni a fornire i propri contenuti a Europeana accrescendo la loro visibilità[17]: oltre alle iniziative già annoverate, è stata indetta una petizione online che vuole sensibilizzare il grande pubblico sulla necessità di tutelare e valorizzare la cultura del cibo e del bere ribadendo l'importanza di questa tematica per la comunità, la cultura e l'economia[18].

Anche l'ICCU ha dato vita, in cooperazione con l'Associazione Wikimedia Italia e la Biblioteca Universitaria Alessandrina, a una gara organizzata nella più ampia cornice del 3° concorso di rilettura del progetto Wikisource. I partecipanti sono invitati a rivedere e curare la formattazione di tre testi, in formato digitale, forniti dalla Biblioteca Universitaria Alessandrina e che, grazie al lavoro svolto nel corso della competizione, saranno più leggibili, stampabili e navigabili in maniera ipertestuale.

Le opere scelte trattano, appunto, argomenti legati all'enogastronomia e sono: l'*Epulario* di Giovanni Rosselli, *Il rimedio infallibile che conserva le quarantine d'anni il vino* di Giovanni Antonio Fineo e il *Trattato della neve e del bere fresco*, traduzione italiana a cura di Giovan Battista Scarampo della sezione dedicata al bere fresco dell'opera *Historia Medicinal*, redatta da Nicolas Monardes.

I premi, per questo concorso chiamato "A banchetto con duchi e papi: un concorso per

TeaTrailLondon banner

[16] http://foodanddrinkeurope.eu/applications/elearning-resources/europeana-attribution-tool/. Per maggiori informazioni sul funzionamento dello strumento, <http://foodanddrinkeurope.eu/applications/elearning-resources/europeana-attribution-tool/>.

[17] http://foodanddrinkeurope.eu/get-involved/become-a-content-provider/.

[18] http://foodanddrinkeurope.eu/get-involved/become-a-supporter/.

SISTEMA MODUS

rileggere antichi ricettari e libri di gastronomia", sono offerti da Wikimedia Italia. La competizione, iniziata il 23 novembre 2015, terminerà l'8 dicembre[19].
Non resta altro che segnalare tutti i mezzi che consentano di restare aggiornati sul progetto: oltre al già citato sito Web (aggiornato recentemente), si segnalano il blog[20], la newsletter[21] e i canali social Twitter, Facebook, Pinterest e Instagram[22]. Sono, dunque, molteplici le possibilità di seguire da vicino l'evoluzione di Food and Drink e tutte le attività in corso e future.

[19] Per informazioni sul concorso, <https://it.wikisource.org/wiki/Wikisource:Dodicesimo_compleanno_di_Wikisource>. Per il comunicato stampa, <http://www.iccu.sbn.it/opencms/opencms/it/archivionovita/2015/novita_0029.html>.
[20] http://foodanddrinkeurope.eu/news/blog/.
[21] Per sottoscrivere e leggere le newsletter, <http://foodanddrinkeurope.eu/news/newsletters/>.
[22] Twitter, <https://twitter.com/foodanddrink_eu>;
Facebook, <https://www.facebook.com/FoodAndDrink.EU>;
Instagram, <https://www.instagram.com/foodanddrinkeurope/>;
Pinterest, <https://www.pinterest.com/FoodAndDrinkEU/>.

L'ultima consultazione dei siti Web è avvenuta nel mese di dicembre 2015.

ICCU

Pubblicazioni

Futuro delle memorie digitali e patrimonio culturale, a cura di Vittoria Tola
e Cecilia Castellani (2004)
ISBN 978-88-7107-110-7

Memorie digitali: rischi ed emergenze, a cura di Alessandra Ruggiero e Vittoria Tola (2005)
ISBN 978-88-7107-111-4

Linee guida per registrazioni d'autorità e di rinvio (2005)
ISBN 978-88-7107-105-3

Linee guida per la digitalizzazione del materiale fotografico, a cura del Gruppo di lavoro
sulla digitalizzazione del materiale fotografico (2005)
ISBN 978-88-7107-112-1

EAD: Descrizione archivistica codificata: dizionario dei marcatori, a cura di Giovanni Michetti (2005)
ISBN 978-88-7107-115-2

MAG: metadati amministrativi e gestionali: manuale utente, a cura di Elena Pierazzo (2006)
ISBN 978-88-7107-117-6

Linee guida per la digitalizzazione del materiale cartografico, a cura del Gruppo di lavoro
sulla digitalizzazione del materiale cartografico (2006)
ISBN 978-88-7107-116-9

Inter omnes: contributo allo studio delle marche dei tipografi e degli editori
italiani del XVI secolo (2006)
ISBN 978-88-7107-114-5

Le edizioni italiane del XVI secolo: censimento nazionale, Vol. 5: D (2006)
ISBN 978-88-7107-113-8

Linee guida per la digitalizzazione di bandi, manifesti e fogli volanti, a cura del Gruppo di lavoro
sulla digitalizzazione di bandi, manifesti e fogli volanti (2006)
ISBN 978-88-7107-118-3

ISBD(CR): International Standard Bibliographic Description for Serials and
Other Continuing Resources: revisione dell'ISBD(S) (2007)
ISBN 978-88-7107-029-2

Miniature e decorazioni dei manoscritti, a cura di Emilia Ambra, Angela Dillon Bussi, Massimo Menna (2006)
ISBN 978-88-7107-119-0

SISTEMA MODUS

Le edizioni italiane del XVI secolo: censimento nazionale, Vol. 6: E-F (2007)
ISBN 978-88-7107-125-1

Il libro italiano del XVI secolo: conferme e novità in EDIT16: atti della giornata di studio, a cura di Rosaria Maria Servello (2007)
ISBN 978-88-7107-124-4

OAIS: Sistema operativo informatico per l'archiviazione, a cura di Giovanni Michetti (2007)
ISBN 978-88-7107-126-8

ISBD(CR): International Standard Bibliographic Description for Serials and Other Continuing Resources: revisione dell'ISBD(S) (2007)
ISBN 978-88-7107-104-6

Regole italiane di catalogazione: REICAT, a cura della Commissione permanente delle regole italiane di catalogazione (2009)
ISBN 978-88-7107-127-5

Requisiti funzionali per i dati di autorità. Un modello concettuale, a cura di Glenn E. Patton. Roma, Iccu (2010)
ISBN 978-88-7107-129-9

IFLA: International Standard Bibliographic Description (ISBD), Roma, ICCU (2010)
ISBN 978-88-7107-128-2

Guida alla catalogazione SBN. Musica: musica e libretti a stampa. Registrazioni sonore, video e risorse elettroniche musicali, Roma, ICCU (2012)
ISBN 978-88-7107-130-5

ISBD: International Standard Bibliographic Description, Edizione consolidata, Roma, ICCU (2012)
ISBN 978-88-7107-131-2

Titolo Uniforme Musicale: Norme per la redazione Roma, ICCU (2014)
ISBN 978-88-7107-133-6

Per l'acquisto delle edizioni ICCU rivolgersi a:

ICCU Ufficio vendite
Viale Castro Pretorio, 105
00185 Roma
T +39 06 49210430
F +39 06 4959302
e-mail: venditapubbl@iccu.sbn.it
http://www.iccu.sbn.it

SISTEMA MODUS

ICCU

Istituto centrale per il catalogo unico
delle biblioteche italiane e per le informazioni bibliografiche
http://www.iccu.sbn.it

Copyright © ICCU - Roma

Finito di stampare nel mese di dicembre 2015
a cura dell'Istituto Poligrafico e Zecca dello Stato S.p.A. - Roma

SISTEMA MODUS

Digitalia

ICCU

Istituto centrale per il catalogo unico
delle biblioteche italiane e per le informazioni bibliografiche
http://www.iccu.sbn.it

Anno X, Numero 1/2 - **2015**

SISTEMA MODUS